Technische Universität Berlin

# Hybrid Analog Digital Beamforming: Implementation and Applications

vorgelegt von

Dipl.-Ing.

Thomas Kühne

ORCID: 0000-0002-8297-4022

an der Fakultät IV - Elektrotechnik und Informatik

der Technischen Universität Berlin

zur Erlangung des akademischen Grades

**Doktor der Ingenieurwissenschaften**

**- Dr.-Ing. -**

genehmigte Dissertation

Promotionsausschuss:

| | |
|---|---|
| Vorsitzender: | Prof. Dr.-Ing. Thomas Magedanz |
| Gutachter: | Prof. Giuseppe Caire, Ph.D. |
| Gutachter: | Prof. Dr.-Ing. Eckhard Grass |
| Gutachter: | Prof. Konstantinos Psounis, Ph.D. |

Tag der wissenschaftlichen Aussprache: 29. April 2022

Berlin 2022

# Abstract

In recent years, two of the main trends in wireless communication have been higher carrier frequencies and larger antenna arrays. Each trend by itself allows for a significant gain in system performance, but they are difficult to combine. The former trend, higher carrier frequencies in the millimeter-wave (mm-wave) range, offers an unprecedented available bandwidth, but it also requires advanced and more complex hardware technology. The latter trend consists in using a large number of antenna elements together with multi-user multiple-input multiple-output (MU-MIMO) processing. Combining these two trends is difficult because systems with many antennas operating in the mm-wave range quickly become too complex. Many publications suggest using the hybrid digital-analog (HDA) beamforming architecture for mm-wave MU-MIMO communication systems, but few comprehensively consider all aspects of the HDA architecture and its implementation.

This dissertation aims to cover all aspects of the HDA architecture, including the hardware, the communication system design, and the signal processing. We examine the interdependencies of all the parts. We introduce the HDA-specific signal processing, present selected algorithms, and describe the hardware of the analog beamforming network. Our investigation suggests both a structure for the analog network and parameters for its components. The signal processing algorithms and the analog hardware are co-simulated and achieve high performance. Power efficiency and complexity are significant issues for mm-wave systems. An analysis of the power efficiency and the complexity of the HDA architecture shows its advantages compared to the fully-digital architecture. We propose a feasible system design for using the HDA architecture in the mm-wave frequency range. A realistic simulation proves the high performance of the system. As a second use case, we propose that the HDA architecture can be used to increase the MU-MIMO capabilities of Wi-Fi systems. This is a solution to increasing Wi-Fi demands that does not require any change in the Wi-Fi protocol. We examined this use case by building a fully functional demonstrator based on commercial off-the-shelf (COTS) Wi-Fi hardware and a self-developed analog beamforming module. We validated the use case by taking measurements with the demonstrator. Using COTS user stations, we measured an increase of 50 % in the end-to-end system throughput. In conclusion, we show that the HDA architecture provides excellent performance and reduces complexity. The hybrid concept can enable the success of future mm-wave systems.

# Zusammenfassung

Zwei der wichtigsten Trends in der drahtlosen Kommunikation in den letzten Jahren waren sehr hohe Trägerfrequenzen und große Gruppenantennen. Der Millimeterwellen-Frequenzbereich (mm-Wellen-Frequenzbereich) bietet mit der großen verfügbaren Bandbreite die Möglichkeit, die Systemleistung zu steigern, erfordert aber auch eine fortschrittlichere und komplexere Hardwaretechnologie. Eine zusätzliche signifikante Steigerung der Systemleistung ist möglich, wenn eine große Anzahl von Antennenelementen zusammen mit einer Mehrnutzer-Mehrantennen-Verarbeitung (MU-MIMO-Verarbeitung, englisch: Multi-User-Multiple-Input-Multiple-Output) verwendet wird. Viele Veröffentlichungen schlagen die Verwendung der hybriden digital-analogen (HDA) Strahlformungsarchitektur für mm-Wellen-MU-MIMO-Kommunikationssysteme vor, aber nur wenige berücksichtigen dabei zusammenhängend alle Aspekte der HDA-Architektur und ihrer Implementierung.

Ziel dieser Dissertation war es, alle Aspekte der HDA-Architektur abzudecken, z. B. die Hardware, das Design des Kommunikationssystems und die Signalverarbeitung. Dabei wurden auch die Abhängigkeiten zwischen den Teilen untersucht. Wir stellen die HDA-spezifische Signalverarbeitung vor, präsentieren ausgewählte Algorithmen und beschreiben die Hardware des analogen Stahlformungsnetzwerks. Unsere Untersuchung schlägt sowohl eine Struktur für das analoge Netzwerk als auch Parameter für seine Komponenten vor. Die Signalverarbeitungsalgorithmen und die analoge Hardware werden gemeinsam simuliert und erreichen eine hohe Systemleistung. Eine Analyse der Leistungseffizienz und der Komplexität der HDA-Architektur zeigt ihre Vorteile im Vergleich zu einer volldigitalen Architektur. Wir schlagen ein umsetzbares Systemdesign für die Anwendung im mm-Wellen-Frequenzbereich vor. Eine realistische Simulation beweist die hohe Leistungsfähigkeit. Als zweiter Anwendungsfall wird die HDA-Architektur als Lösung vorgeschlagen, um die MU-MIMO-Leistung von Wi-Fi-Systemen zu erhöhen. Unsere Lösung erfordert keine Änderung des Wi-Fi-Protokolls. Wir haben einen voll funktionsfähigen Demonstrator gebaut, der auf handelsüblicher Wi-Fi-Hardware und einem selbst entwickelten analogen Strahlformungsmodul basiert. Unser Ansatz wird durch Messungen mit dem Demonstrator validiert. Wir messen mit handelsüblichen Nutzergeräten eine Steigerung des Ende-zu-Ende-Durchsatzes um 50 %. Zusammenfassend lässt sich sagen, dass die HDA-Architektur eine hervorragende Systemleistung bietet und die Komplexität reduziert. Das hybride Konzept kann den Erfolg zukünftiger mm-Wellensysteme ermöglichen.

# Acknowledgements

I like to thank my advisor Prof. Giuseppe Caire for giving me the opportunity of doing a Ph.D. and for supporting me over the years. Thanks to his continued support and guidance, I learned a lot during the last few years. I enjoyed our discussions on the latest trends in mobile communication, during which he inspired me by his vast knowledge and interest. I would also like to thank Prof. Konstantinos Psounis and Prof. Eckhard Grass for agreeing to review this work and being part of the committee. The Alexander-von-Humboldt Foundation and the European Union financially supported my work. A part of this thesis is related to the SERENA project, which is funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 779305. It is great how the European Union enables research cooperation with many partners from different countries. This shaped my professional collaboration and intercultural skills. I also want to thank all of the SERENA project partners for this experience.

Many thanks go to the entire team of the Communications and Information Theory Group. I would like to especially thank Jana for her help with the administrative work and for always having an open ear for all non-scientific problems. I would also like to thank Andreas Kortke, Xiaoshen Song, Piotr Gawłowicz for their collaboration and our fruitful scientific discussions. Without their work, this thesis would certainly be less interesting. In addition, I would like to thank Mahdi for proofreading the first part of this thesis and Alexander for sharing his thesis template. I thank the rest of the group, they are too many to name them all. It was a pleasure to work with everyone in the group and, the friendly atmosphere was always a motivation and support. I have fond memories of the enriching and inspiring conversations during the lunch and coffee breaks.

Ich möchte auch meinen Eltern und meiner Schwester für ihre Unterstützung danken. Ihre Ermunterung, stets neugierig zu sein und neues Wissen zu erlernen, hat mich erst auf den Weg zur Promotion gebracht. Ein besonderer Dank geht an meine geliebte Rebecca. Gerade am Ende der Promotion hat deine Unterstützung mich vorangebracht und ohne dich wäre ich nicht die Person, die ich heute bin.

*Thomas Kühne,*
*Berlin, January 2022*

# Contents

# List of Figures

# List of Tables

# List of Abbreviations

| | | |
|---|---|---|
| **5G** | fifth-generation standard for cellular networks | 2 |
| **ADC** | analog-to-digital converter | 12 |
| **AM** | analog module | 69 |
| **AoA** | angle of arrival | 3 |
| **AoD** | angle of departure | 3 |
| **AP** | access point | 1 |
| **ASIC** | application-specific integrated circuit | 64 |
| **BA** | beam alignment | 3 |
| **BBF** | before beamforming | 88 |
| **BS** | base station | 1 |
| **BSS** | basic service set | 107 |
| **BST** | beam steering | 35 |
| **BZF** | baseband zero-forcing | 36 |
| **CDF** | cumulative distribution function | 109 |
| **COTS** | commercial off-the-shelf | 5 |
| **CSI** | channel state information | 3 |
| **DAC** | digital-to-analog converter | 12 |
| **FC** | fully-connected | 3 |
| **FDD** | frequency-division duplexing | 10 |
| **FPGA** | field-programmable gate array | 64 |
| **HDA** | hybrid digital-analog | 2 |
| **IC** | integrated circuit | 22 |
| **JSDM** | joint spatial division and multiplexing | 13 |
| **L-LTF** | legacy long training field | 107 |

**LNA** low-noise amplifier . . . 55
**LOS** line-of-sight . . . 30
**MCS** modulation and coding scheme . . . 101
**MIMO** multiple-input multiple-output
**mm-wave** millimeter-wave . . . 2
**MU-MIMO** multi-user multiple-input multiple-output . . . 1
**NDP** null data packet . . . 106
**NF** noise figure . . . 55
**NIC** network interface card . . . 97
**NLOS** non-line-of-sight . . . 30
**OFDM** orthogonal frequency division multiplexing . . . 27
**OSPS** one-stream-per-subarray . . . 3
**PA** power amplifier . . . 54
**PAPR** peak-to-average power ratio . . . 57
**PRACH** physical random access channel . . . 86
**RF** radio frequency . . . 2
**RMS** root mean square . . . 75
**RSSI** received signal strength indicator . . . 100
**RX** reception . . . 54
**SC** subcarrier . . . 36
**SCM** single-carrier modulation . . . 57
**SDR** software-defined radio . . . 67
**SINR** signal-to-interference-plus-noise ratio . . . 103
**SNR** signal-to-noise ratio . . . 11
**SSB** synchronization signal block . . . 85
**STA** user station . . . 6
**TX** transmission . . . 54
**TDD** time-division duplexing . . . 10
**UE** user equipment . . . 6

**ULA** uniform linear array . . . . . 20

**URA** uniform rectangular array . . . . . 20

**VHT-LTF** very high throughput long training field . . . . . 107

**VNA** vector network analyzer . . . . . 73

**WLAN** wireless local area network . . . . . 1

**ZF** zero-forcing . . . . . 10

# 1. Introduction

Wireless communication is ubiquitous in our modern lives. The wireless data traffic has seen a dramatic increase in recent years and the massive growth is predicted to continue. Ericsson reports a growth of the global mobile data traffic of above 50 % per year between 2014 and 2021 and estimates at least a growth of 30 % per year until 2026 [1]. This was and will be possible due to the increased performance of wireless communication technologies. According to Cisco, the average data rates of both mobile cellular networks and wireless local area networks (WLANs) will triple between 2018 and 2023 [2]. This progress is driven by innovative research. One key technology leap to increase the data rates of both communication network types was the introduction of the multi-user multiple-input multiple-output (MU-MIMO) technology. With MU-MIMO, base stations (BSs) in cellular networks and access points (APs) in WLANs can communicate with multiple users simultaneously at the same time in the same frequency band [3], [4]. Several data streams are multiplexed in the spatial domain to different users (i.e., users are at different locations in space). This is achieved using multiple antenna elements, usually arranged in an antenna array, at the BS or AP. In the downlink, the data streams are processed by a MU-MIMO precoder and transmitted via the antenna array. In general, the gain of MU-MIMO precoding depends, among other things, on the wireless channel and the precoding algorithm. Interference between the user channels limits the system spatial multiplexing gain and achievable downlink rate. Except in special cases, a multi-user channel always introduces interference between the users. In the downlink, the data streams are processed by a MU-MIMO precoder whose purpose is to mitigate inter-user interference and to maximize a performance metric such as the downlink data rate [4].

MU-MIMO was first introduced to WLANs in 2013 with the IEEE 802.11ac revision of the Wi-Fi standard [5], [6]. The newly released version IEEE 802.11ax extends the MU-MIMO capabilities even further. Wi-Fi now supports up to eight spatial data streams in both downlink and uplink [7]. Unfortunately, the commercial adoption of MU-MIMO for Wi-Fi networks has been so far quite slow. Current commercially available high-end APs have commonly four antenna elements for each supported frequency band and support up to three concurrent users in each frequency band [8]. Hence, Wi-Fi networks do not yet utilize the full advantage of the MU-MIMO operation as defined in the 2013 standard.

In cellular networks, MU-MIMO was first introduced in the fourth generation standard (4G) in 2008 [9]. In the following years, the technology in cellular networks evolved and the capabilities of MU-MIMO were greatly extended. In fact, one of the key performance drivers of the fifth-generation standard for cellular networks (5G) is massive MIMO [10], a MU-MIMO technology. In Massive MIMO, the number of BS antenna elements is much larger than the number of user devices. As a consequence, massive MIMO systems can achieve very high MU-MIMO gains [11]. The first commercial 5G massive MIMO systems can already serve up to eight user devices simultaneously using spatial multiplexing [12]. Researchers even demonstrated a 5G MU-MIMO testbed achieving a 22-fold MU-MIMO gain with respect to the single-antenna reference system [13].

Massive MIMO can be implemented using different precoding/combining algorithms and even different signal processing architectures [4]. The hybrid digital-analog (HDA) architecture is one of the concepts to implement massive MIMO. In an HDA system, the MU-MIMO signal processing at the BS/AP is divided into a low-dimensional digital precoding and a high-dimensional analog beamforming [14]. The analog part employs a controllable hardware network to connect a large number of antenna elements to a small number of radio frequency (RF) chains. In short, an RF chain converts an analog signal to its digital representation and vice versa. The reduced number of RF chains relative to the number of antenna elements decreases the hardware complexity [14]. In addition, due to the small number of digital signals, the digital processing is less complex compared to other massive MIMO architectures. The HDA architecture is seen as a solution to hardware complexity, power consumption, and channel estimation issues that arise in massive MIMO systems [10], [14].

Apart from massive MIMO, the shift towards the millimeter-wave (mm-wave) frequency range is a key performance driver of 5G [10], [15]. Current mm-wave systems operate in the frequency range above 6 GHz and below 100 GHz. The main reason for the use of higher frequencies is the availability of unused or lightly used bandwidth [15]. The new mm-wave frequency range 2 (FR2) of the 5G NR standard lists a total available bandwidth of more than 12 GHz [16]. Both the wireless channel and the system design at mm-wave bands have distinctive characteristics and were widely investigated during the last years [10], [15], [17]–[21]. The thesis [21] gives a comprehensive introduction to the mm-wave channel characteristics.

Although the use of the HDA architecture is not limited to any specific frequency band, the popularity of mm-wave 5G systems has resulted in a particular interest in the application of the HDA architecture to mm-wave systems (see, e.g., the survey [14]). The HDA architecture has various specific implementation aspects. The following list gives an

overview of the main topics with respect to signal processing and communication system design, and includes references to the state of the art.

- *Analog network.* The analog beamforming network can have different structures also known as "architectures" in the literature. The structure defines the way how the antenna elements are connected to the RF chains. The two most commonly considered structures are the fully-connected (FC) structure and the sub-connected or one-stream-per-subarray (OSPS) structure [14], [17], [22]–[25]. The former outperforms the latter at the cost of complexity [23], [25]. Besides the structure, the choice of the controllable elements can vary. [22], [23].

- *Power efficiency.* As power consumption is one of the reasons to use the HDA architecture, different works consider its power efficiency. The power efficiency depends, e.g., on the structure of the analog network, and the location and the type of the used amplifiers [17], [24], [26], [27].

- *Channel estimation techniques.* The structure of the HDA architecture restricts the simple measurement of the channel state information (CSI). Not all antenna elements can be simultaneously measured at the same time. The channel or some representation (e.g., the angle of arrival (AoA)/angle of departure (AoD) pairs) is estimated during the initial access by special algorithms [17], [28]–[31]. We refer to the initial access phase and the estimation as beam alignment (BA).

- *Precoding/combining algorithms.* The MU-MIMO precoding needs to be adapted to the split into the digital and the analog part. The limitations of the analog network influence the precoding design and the algorithms [18], [19], [27], [32], [33]. The precoder calculation requires the CSI. The CSI is estimated during the BA phase and can be different compared to a standard MU-MIMO system [27].

- *Experimentation and prototypes.* All of the previously mentioned literature relies on calculations and simulations. Simulations are often based on assumptions. Experimentation can validate the assumptions and the results. Experimentation enables the transition of the HDA technology towards commercial products. The publication [34] gives an overview of prototypes and testbeds of mm-wave systems.

Most of the literature listed above does only consider a single aspect or assumes idealized hardware. This thesis tries to combine realistic hardware considerations, well-founded choices for the signal processing algorithms, and experimentation. Besides the application in mm-wave cellular networks, the thesis also includes an investigation on the application for WLANs.

## 1.1. Outline and Contributions

As mentioned above, this dissertation addresses the main aspects of the HDA architecture and their interdependencies. It attempts to cover all HDA-specific parts necessary to implement a system. The goal is to offer insights for designing and implementing HDA systems for wireless communication. An outline of the thesis is given below.

The dissertation consists of two parts. The first part describes the motivation, the general concept, and the distinctive components of an HDA system. It starts with Chapter 2, explaining the background and the motivation underlying the use of the HDA structure. We introduce the concept of MU-MIMO precoding and some properties necessary to evaluate different precoding systems. We show how massive MIMO can achieve a very high spatial multiplexing gain. We explain the difference between the fully-digital massive MIMO architecture and the HDA architecture. From this comparison, we conclude for which application scenarios the HDA architecture is recommended. In Chapter 3, we discuss the implementation details of the HDA architecture. We describe the specific parts of an HDA system and how they interact. An integral part of an HDA wireless communication system is the signal processing which needs to be adapted to the hybrid architecture. We describe the initial access phase with the BA process and the data communication phase with the MU-MIMO precoding. For both problems, we introduce an algorithm proposed in the literature and review its advantages and disadvantages. Chapter 3 also includes a section on the implementation details and considerations related to the hardware. The goal of this section is to discuss certain parameters, features, and limitations of typical hardware implementations, and not just the properties of a state-of-the-art implementation. The main variants of the structure of the analog beamforming network are explained. The structures are compared concerning their performance, limitations, and requirements. We then describe the main components of the network which are used to control the phase and amplitude of the network paths. Their properties, e.g., the resolution and the non-ideal characteristics, are investigated. As amplification is an important part of every RF system, we discuss its relationship to the HDA architecture. Most considerations, e.g., the location of the amplifiers within the network, are not strict but design choices. We try to outline the possibilities and give recommendations. Closely related to the amplification but also the structure of the network is the power efficiency. We investigate the power efficiency of the two main structures. Following the power efficiency, we analyze the complexity of HDA systems. We discuss the conceptual differences in complexity between the HDA architecture and the standard fully-digital massive MIMO architecture. At the end of Chapter 3, a short section addresses the digital signal processing hardware. The digital hardware in an HDA system is not specific but similar to other MU-MIMO systems. Thus, we only present an

overview of some aspects. Following the abstract discussion, we describe our self-developed analog beamforming module in Chapter 4. The module was developed for experimentation and used to achieve the results presented in the second half of the thesis.

The second half of the thesis investigates actual HDA systems adapted to their application scenario. It presents systems for two applications, one for cellular networks and one for WLANs. The application for mm-wave cellular networks was the main driver of the research on the HDA architecture. Chapter 5 presents a system for this application. The system, which is developed with many partners in the European research project SERENA[1], is intended for the mm-wave frequency range of 5G. We describe the overall system and certain implementation choices. This includes sections on the antenna array, analog beamforming network, and signal processing. We study the requirements on the signal processing from the 5G standard. Unfortunately, the system is not yet in operation and, hence, we can only present simulation results of different performance measures. We use a geometry-based stochastic channel simulator and simulate both certain hardware aspects and signal processing parts. Chapter 6 covers the application for WLANs. As mentioned in the introduction, the MU-MIMO capabilities of current Wi-Fi networks are limited. We propose the HDA architecture as a solution. Specifically, we enable the Wi-Fi access point to reach its full MU-MIMO capability. Our solution does not require any change in the Wi-Fi protocol. We present a demonstrator based on commercial off-the-shelf (COTS) Wi-Fi hardware, the analog beamforming module (see Chapter 4), and a novel control protocol. The implemented control protocol and signal processing are fully transparent to the Wi-Fi part. We evaluate the concept and our demonstrator through measurements. We first show the positive effect of our concept on the wireless channel seen by the Wi-Fi system by measuring the condition number of the channel. Second, we measure with COTS Wi-Fi user stations an increase in the end-to-end system throughput of 50 %. Chapter 7 concludes the thesis and provides suggestions for future work.

## 1.2. Notation

In this thesis, scalars are denoted by non-boldface letters (e.g, $x$, $X$), vectors by small boldface letters (e.g, $\mathbf{x}$), matrices by capital boldface letters (e.g, $\mathbf{X}$), and sets by calligraphic letters (e.g., $\mathcal{X}$). Unless otherwise defined, the $i$-th element of a vector $\mathbf{x}$ is represented by $[\mathbf{x}]_i$ and the $i,j$-th element of a matrix $\mathbf{X}$ by $[\mathbf{X}]_{i,j}$. For a vector $\mathbf{x}$, the symbol $\mathrm{diag}(\mathbf{x})$ denotes a matrix with $\mathbf{x}$ as its main diagonal. For multiple vectors or multiple matrices, the symbol $\mathrm{diag}(\mathbf{x}_1, \ldots, \mathbf{x}_k)$ or $\mathrm{diag}(\mathbf{X}_1, \ldots, \mathbf{X}_k)$ denotes a block diagonal matrix with

[1]The SERENA project was funded from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 779305. More information on `https://serena-h2020.eu/`.

$\mathbf{x}_1, \ldots, \mathbf{x}_k$ or $\mathbf{X}_1, \ldots, \mathbf{X}_k$ aligned on the diagonal. The symbol $\mathbf{I}_K$ denotes the identity matrix of size $K \times K$. The complex conjugate, transpose, conjugate transpose, and inverse are represented by the superscripts $(\cdot)^*$, $(\cdot)^\mathsf{T}$, $(\cdot)^\mathsf{H}$, and $(\cdot)^{-1}$, respectively. The expectation of $X$ is $\mathbb{E}[X]$. For an integer $K \in \mathbb{Z}$, we use the shorthand notation $[K]$ for the index set $\{1, \ldots, K\}$. The phase $\phi$ and the absolute value $r$ of the complex number $z = r\, e^{i\phi}$ are denoted as $\arg(z)$ and $|z|$, respectively.

The term millimeter-wave (mm-wave) is used throughout this work to describe a certain frequency range. Signals with a wavelength between 1 mm and 10 mm, so an approximate frequency between 30 GHz and 300 GHz, are in the mm-wave range. In most literature, both scientific and non-scientific, the term is used for signals with a frequency above 6 GHz. This mainly originated from research on the fifth-generation standard for cellular networks (5G) where frequency bands above 6 GHz were used for the first time on a large scale in mobile communication. This work follows the same common definition of the term. In contrast, the term sub-6 GHz is used in this thesis to describe the frequency range below 6 GHz.

The terms HDA, hybrid beamforming, or only hybrid (e.g., in hybrid concept) all refer to the same concept in this thesis. In most wireless networks, communication happens between a central node and user devices. We call the central node in cellular networks base station (BS) and in WLANs access point (AP). The term user equipment (UE) is usually used for a user device in cellular networks and the term user station (STA) in WLANs. In Chapters 1 to 4, we will use BS/AP for the central node and UE for the user devices.

## 1.3. Publications and Copyright Disclaimer

During my Ph.D. studies, I have authored or co-authored multiple publications and research reports. I have re-used in this work some material from these previously published or submitted journal and conference papers and the research reports. The following list includes all my publications that are related to this thesis.

1. T. Kühne and G. Caire, "An Analog Module for Hybrid Massive MIMO Testbeds Demonstrating Beam Alignment Algorithms", in *22nd International ITG Workshop on Smart Antennas (WSA 2018)*, Bochum, Germany, Mar. 2018

2. T. Kühne, G. Caire, and X. Song, "Signal processing algorithms and specifications", SERENA, research rep., Jan. 2018. DOI: 10.5281/zenodo.3240455

3. X. Song, T. Kühne, and G. Caire, "Fully-Connected vs. Sub-Connected Hybrid Precoding Architectures for mmWave MU-MIMO", in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, May 2019. DOI: 10.1109/ICC.2019.8761521

4. K. Andersson and T. Kühne, "Proof of Concept Platform and Front end Specifications", SERENA, research rep., Jun. 2019. DOI: 10.5281/zenodo.3240304

5. X. Song, T. Kühne, and G. Caire, "Fully-/Partially-Connected Hybrid Beamforming Architectures for mmWave MU-MIMO", *IEEE Transactions on Wireless Communications*, vol. 19, no. 3, pp. 1754–1769, Mar. 2020, ISSN: 1536-1276. DOI: 10.1109/twc.2019.2957227

6. T. Kuehne, X. Song, G. Caire, *et al.*, "Performance Simulation of a 5G Hybrid Beamforming Millimeter-Wave System", in *24th International ITG Workshop on Smart Antennas (WSA 2020)*, Hamburg, Germany, Feb. 2020

7. T. Kühne, P. Gawłowicz, A. Zubow, *et al.*, "Demo: Bringing Hybrid Analog-Digital Beamforming to Commercial MU-MIMO WiFi Networks", in *26th ACM International Conference on Mobile Computing and Networking (MobiCom 2020)*, London, United Kingdom, Sep. 2020, ISBN: 978-1-4503-7085-1. DOI: 10.1145/3372224.3417320

8. T. Kühne, P. Gawłowicz, A. Zubow, *et al.*, "Hybrid Analog-Digital Beamforming: Unlocking the Real MU-MIMO Potential of Commodity WiFi", *(to be submitted)*, 2021

# 2. Background and Motivation

The HDA architecture is an approach to implement massive MIMO systems, with advantages compared to other architectures in certain application scenarios. This chapter gives the background to understand when the hybrid architecture should be used. The chapter is partly based on the problem statement in an original journal manuscript [39] which will be submitted to the IEEE/ACM Transactions on Networking. As a simplification, in this chapter, we only consider the downlink traffic from the BS/AP to the UEs.

## 2.1. Multi-User-MIMO Background

We assume that the BS/AP is equipped with $M$ antenna elements and communicates with users with total $N$ antenna elements (e.g., $N$ single-antenna UEs). Therefore the fading channel between the BS/AP and the UEs can be mathematically represented by $\mathbf{H}$ an $N \times M$ matrix which we will refer to as the channel state information (CSI). The maximum number of supported spatial streams, or the maximum achievable MU-MIMO gain, is limited by the rank of $\mathbf{H}$. The maximum of the rank is $\min\{M, N\}$ when $\mathbf{H}$ is said to have full rank. The row $k$ of $\mathbf{H}$ is the channel vector of user $k$ of dimension $M$. In the case of interference between the users, the rows of $\mathbf{H}$ are non-orthogonal vectors and in the extreme case linearly dependent. Although, the rows of $\mathbf{H}$ are only linearly dependent if some users have the exact same set of AoD/AoA. Such a channel is not full-rank and no precoding algorithm can achieve the maximum MU-MIMO gain. Typical channels have an $\mathbf{H}$ with a full rank but still non-orthogonal channel vectors towards the users.

The standard signal model for the downlink transmission from the BS/AP to all $N$ users is given by

$$\mathbf{y} = \mathbf{H} \cdot \hat{\mathbf{x}} + \mathbf{z}, \tag{2.1}$$

where $\hat{\mathbf{x}} \in \mathbb{C}^M$ is the transmitted signal vector over the $M$ antenna elements, $\mathbf{z} \in \mathbb{C}^N$ is the additive white Gaussian noise vector at the receivers, and $\mathbf{y} \in \mathbb{C}^N$ is the vector of the received signals of all $N$ users. Considering linear precoding, the transmit signal is given as

$$\hat{\mathbf{x}} = \mathbf{W} \cdot \mathbf{x}, \tag{2.2}$$

where $\mathbf{W} \in \mathbb{C}^{M \times K}$ is the precoding matrix and $\mathbf{x} \in \mathbb{C}^{K}$ is the vector of the data symbols. In general, $K \leq \min\{M, N\}$ denotes the number of downlink spatial streams transmitted on each channel use, which coincides in this thesis with the number of simultaneously served UEs.[1]

As the precoding algorithm depends on $\mathbf{H}$, the transmitter requires the CSI to calculate the precoder for the downlink. The CSI is acquired by the transmitter either using channel reciprocity (i.e., measuring the CSI in the uplink for the use in the downlink) or some form of feedback from the users [4]. Channel reciprocity is only given in systems that use the same frequency band in the uplink and the downlink (e.g., time-division duplexing (TDD) systems). Frequency-division duplexing (FDD) systems require some form of feedback from the users. In general, the CSI acquisition introduces some kind of overhead, which decreases the total system data rate as it uses resources that could be used for communication. For example, in an FDD system, the feedback would be sent in the uplink in a special message and would require time that can not be used for the uplink data traffic.

If we assume that $\mathbf{H}$ is perfectly known at the transmitter, the interference created by a typical channel (i.e., $\mathbf{H}$ has a full rank) can be perfectly removed by precoding [40], [41]. Perfect channel knowledge can be assumed, in case the CSI measurement and feedback system provides sufficiently good quality CSI knowledge such that the residual interference due to non-perfect knowledge is not the dominant factor in the user rates w.r.t. to the noise floor of the receivers. Even though typical precoding algorithms like zero-forcing (ZF) [40] can perfectly remove the interference, they suffer a penalty. Removing the interference comes at the cost of also attenuating the useful signals. The limited power of the transmitter is split between the power for the interference cancellation and the user signals. The received signal by user $k$ after ZF is given by

$$y_k = \sqrt{\alpha_k} \cdot x_k + z_k, \tag{2.3}$$

where $x_k$ is the transmitted data, $z_k$ is the additive white Gaussian noise at the receiver, and $\alpha_k$ is the residual channel gain after ZF. It is well known (e.g., in [42] or [41]) for ZF that

$$\alpha_k = 1/[(\mathbf{H}\mathbf{H}^{\mathsf{H}})^{-1}]_{k,k}. \tag{2.4}$$

Accordingly, $\alpha_k$ is inverse proportional to the condition number of $\mathbf{H}$. The condition number of $\mathbf{H}$ is the ratio between the maximal singular value and the minimal singular value of $\mathbf{H}$. $\alpha_k$ is close to 1 when the condition number is small, and it can be small when the condition number is large. A small $\alpha_k$ means that the received power is reduced

[1] Thus, commonly $K = N$ with $N \leq M$.

by the zero-forcing, which is a power penalty for removing the interference. A channel resulting in small $\alpha_k$ coefficients for some or all users is called ill-conditioned. The power penalty does reduce the signal-to-noise ratio (SNR) for the data stream and, hence, the data rate. As a reference, in [43, Ch. 7.1.4], the authors have explained the same fact by a different approach. They introduce the channel condition number as a measure of favorable propagation for MU-MIMO precoding.

In general, not only ZF but also other precoding algorithms like regularized zero-forcing [44] or dirty paper coding (DPC) [40] incur this penalty. The exact penalty might vary depending on the balance between how much interference is removed and how strong the useful signals are attenuated. $\alpha_k$ for other algorithms might be different to (2.4) but still dependent on the quality of the channel.

To understand how the performance of a MU-MIMO system can be increased, we examine the capacity $C$ of the system. In accordance with the block-fading model, we assume the channel $\mathbf{H}$ to be constant over a time-frequency block of size $T$. Then, the capacity of a MU-MIMO system in the high-SNR regime behaves at best as

$$C(\mathrm{SNR}) = M^*(1 - M^*/T) \log \mathrm{SNR} + O(1), \tag{2.5}$$

where $M^* = \min\{M, N, T/2\}$ [45]. The pre-log factor $M^*(1 - M^*/T)$ is the best possible spatial multiplexing gain or MU-MIMO gain. If the goal is to increase the capacity $C$ in the high SNR regime, the most sensible way is to increase the pre-log factor $M^*$. To increase $M^*$, we have to increase the number of BS/AP antenna elements $M$ and/or the total number of user antenna elements $N$ (until $\min\{M, N\} = T/2$). Systems typically do not always operate in the high-SNR regime but also the intermediate and sometimes the low-SNR regime. The capacity in (2.5) is only proven in the high-SNR regime. In the high-SNR regime, a variation of the SNR can be neglected as the SNR is very large and in the limit $\mathrm{SNR} \to \infty$. In contrast, in the intermediate-SNR regime, a variation of the SNR has an impact on the data rate. Please refer to [43, Ch. 3.4] for the relation between the SNR and a capacity lower bound. Following, not only an increasing pre-log factor (i.e., the number of antennas) but also an increasing SNR (i.e., the received power) can increase the rate. The received power depends on the power penalty $\alpha_k$ of the precoder. As written above, $\alpha_k$ or equivalent the SNR is increased when the channel condition number is small. A channel $\mathbf{H}$ has a small condition number if the channel vectors towards the users are as orthogonal as possible.[2]

[2]The condition number is 1 if the channel vectors towards the users are orthogonal.

## 2.2. Massive MIMO to Increase the MU-MIMO Gain

Massive MIMO was originally proposed to maximize the multi-user sum rate in cellular networks by employing a large number of antennas at the BS relative to the number of users $M \gg N$ [11], [46], [47]. With sufficient users within the cell, the system would achieve an optimal pre-log factor and serve $N = T/2$ users within one coherence block of size $T$. Besides, massive MIMO results in optimal precoding as the channel vectors towards the users become asymptotically orthogonal as $M$ is increasing [11], [47]. Following, even for a realistic (but still much larger than $N$) number of $M$, the channel becomes nearly orthogonal. The condition number of $\mathbf{H}$ becomes close to 1 and the power penalty of the MU-MIMO precoding vanishes. As a result, the sum rate of the system is maximized because both the pre-log factor is maximized and the MU-MIMO precoding does not incur a power penalty. Other significant benefits of massive MIMO are increased energy efficiency (with respect to the transmitted energy) and a simplification in the user scheduling [47].

These benefits come at a cost, the complexity of the hardware increases and is significantly larger than in previous systems. This is true for the RF and the digital signal processing hardware [48]. The original proposal of massive MIMO is based on the so-called fully-digital architecture. In the fully-digital architecture, each antenna element is connected to an RF chain [47]. Hence, each signal of every antenna element has a digital representation. The MU-MIMO signal processing (e.g., the precoding) is done in digital hardware. An RF chain typically consists of amplifiers, filters, RF switches, a frequency up- and down-converter, a local oscillator, an analog-to-digital converter (ADC), and a digital-to-analog converter (DAC). The large number of antenna elements $M$ requires a large number of these components. Depending on the specific design and system parameters, this can increase both the cost and the power consumption of a massive MIMO system compared to previous MU-MIMO systems (where $M$ is not much larger than $N$). In the digital part, the hardware needs to process $M$ data streams at the same time. This requires more processing power and often also high data rates in the digital interconnection links [48].

One assumption for the precoding and the sum rate we made above was the CSI knowledge at the transmitter. When $M$ and $N$ increase, also the size of $\mathbf{H}$ increases. The overhead to acquire the resulting large amount of CSI is an additional cost of massive MIMO. The initial works on massive MIMO proposed to use TDD and channel reciprocity for the CSI acquisition [11], [46]. In a TDD system, the CSI is estimated during the uplink at the BS/AP and used in the same coherence block to compute the precoding for the downlink. Following, no feedback of the CSI is required and the transmitter has always up-to-date instantaneous knowledge of the channel. Since the channel is measured using the $M$ BS/AP antenna elements, the training dimension for the channel acquisition is determined by $N$.

Hence, the training cost for a TDD fully-digital massive MIMO system is acceptable[11]. In contrast, a system based on FDD can not use channel reciprocity to estimate the CSI. The CSI is estimated at the UEs and fed back to the BS/AP. This results in three problems (as pointed out in [41] or [45]). First, the training dimension is determined by $M$ which is very large in a massive MIMO system. Thus, the training overhead during the downlink is large and can not be neglected [41]. Second, since the size of the CSI is large, the feedback overhead in the uplink reduces the uplink sum rate of the system. And third, the feedback might arrive with a delay at the BS/AP. The assumption of the perfect channel knowledge is broken if the delay is larger than the coherence time. The CSI would be outdated at the BS/AP which can be modeled as a prediction error in the sum rate analysis [45]. The prediction error or, in general, the delay reduces the sum rate of the system [49].

Considering that currently deployed cellular networks are often based on FDD systems, it is desirable to solve the CSI acquisition problem. In [50] and [45] the authors have proposed joint spatial division and multiplexing (JSDM) as a solution. The authors have observed that, for typical cellular deployments, the channel vector towards any user is a correlated random vector with a covariance matrix that depends on the channel geometry [45]. In other words, the channel vector towards a user is influenced by two effects, a short-term frequency-dependent and a long-term frequency-flat effect. The short-term effect is modeled as a random process that originated in the movement of the user within its direct scattering environment. The long-term effect is modeled as the covariance matrix of the random process and determined by the channel geometry, e.g., the AoA/AoD pair of the strongest path. The coherence time of the long-term effect is much longer than the coherence time of the short-term effect. The long-term effect is also constant over the channel bandwidth. The coherence time of the short-term effect is equivalent to the coherence time of the overall channel which determined the time-frequency slot size $T$. The approach of JSDM suggests splitting the downlink precoding into two stages [45]. The first stage is a beamforming stage that depends on the channel second-order statistics (i.e., the covariance matrices). The succeeding second stage is a MU-MIMO precoder which depends on the instantaneous channel, inclusive of the beamforming stage. The MU-MIMO precoding matrix $W$ from (2.2) is given as the product

$$\mathbf{W} = \mathbf{B} \cdot \mathbf{P}, \tag{2.6}$$

where $\mathbf{B} \in \mathbb{C}^{M \times b}$ is the beamforming matrix of the first stage, $\mathbf{P} \in \mathbb{C}^{b \times K}$ is the MU-MIMO precoding matrix of the second stage, and $b \leq K$ is a design parameter. The users are grouped according to their channel geometry. In particular, users which are close in space and, hence, have similar AoAs/AoDs are grouped together. The beamforming matrix of the first stage is calculated from the covariance matrices of the groups. Its goal is to minimizes

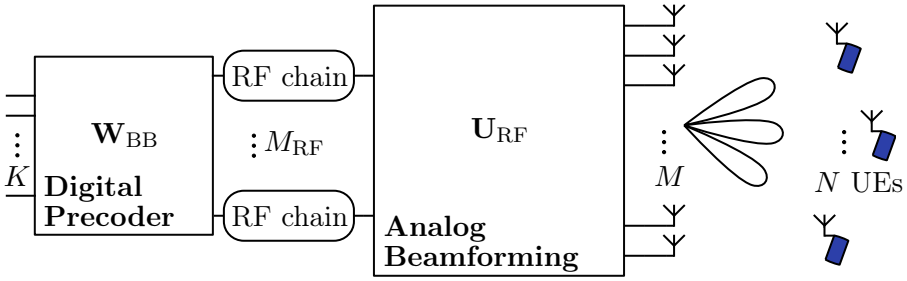



Figure 2.1.: High-level overview of a hybrid system.

the intergroup interference. The MU-MIMO precoder $P$ depends not on the full channel but the effective channel after beamforming. The effective channel matrix $\tilde{\mathbf{H}}$ is given by $\tilde{\mathbf{H}} := \mathbf{H} \cdot \mathbf{B} \in \mathbb{C}^{N \times b}$. The beamforming is designed in a way, that the dimension of the effective channel matrix is reduced compared to the overall channel ($N \times b$ compared to $N \times M$) [45]. The second stage MU-MIMO precoding separates the users within a group and can leverage the reduced intergroup interference to achieve a high MU-MIMO gain.

The training of the CSI for JSDM is also split into two parts. One part measures the covariance information of the overall channel and the other part the instantaneous effective channel. Since the coherence time of the channel covariance information is longer than the coherence time of the overall instantaneous channel, it can be obtained with a much longer time period. Channel covariances are estimated via methods designed to achieve high accuracy with relatively low training overhead [30], [51], [52]. The training for the instantaneous effective channel is done for every coherence block but only for the effective channel. Due to the beamforming and the reduced channel dimension of the effective channel, this training is still feasible [45]. The sum of the overhead of both training parts is much smaller compared to the overhead estimating the full instantaneous channel. Following, JSDM enables FDD massive MIMO systems and makes them a feasible option besides TDD systems.

## 2.3. Hybrid Digital-Analog Architecture for Massive MIMO

The hybrid digital-analog (HDA) architecture is an alternative to the fully-digital signal processing architecture for massive MIMO. As introduced in Chapter 1, HDA systems use both analog beamforming and digital precoding to implement the massive MIMO MU-MIMO processing [14]. Figure 2.1 shows the block diagram of an HDA system. An analog beamforming network connects $M_{\mathrm{RF}}$ RF chains to the $M$ antenna elements. The network consists generally of power dividers/combiners, amplifiers, and vector modulators.

A vector modulator is a combination of a phase shifter and an amplitude modulator. The complexity of the analog network per antenna element is less than the complexity of an RF chain. The analog network is steering beams towards the UEs which is equivalent to forming an effective more diagonally dominant channel for the digital system. Hence, it needs to be adjustable to the user location or rather the current channel. Mathematically, the analog network is represented by a matrix $\mathbf{U}_{\mathrm{RF}} \in \mathbb{C}^{M \times M_{\mathrm{RF}}}$. With $M_{\mathrm{RF}} < M$, the network reduces the channel dimension seen by the digital processing. This effective channel is defined as

$$\tilde{\mathbf{H}} := \mathbf{H} \cdot \mathbf{U}_{\mathrm{RF}} \in \mathbb{C}^{N \times M_{\mathrm{RF}}}. \tag{2.7}$$

The digital processing uses a precoding algorithm similar to the fully-digital MU-MIMO processing. It is represented by $\mathbf{W}_{\mathrm{BB}} \in \mathbb{C}^{M_{\mathrm{RF}} \times K}$, where $K$ is the number of spatial streams. The overall MU-MIMO precoding $\mathbf{W}$ (as used in (2.2)) of the HDA architecture is

$$\mathbf{W} = \mathbf{U}_{\mathrm{RF}} \cdot \mathbf{W}_{\mathrm{BB}}. \tag{2.8}$$

Both $\mathbf{U}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{BB}}$ are optimized depending on the channel $\mathbf{H}$ to maximize the MU-MIMO gain. Following from $M_{\mathrm{RF}} < M$, the maximum spatial gain of an HDA system is $M_{\mathrm{RF}}$, requiring $K \leq M_{\mathrm{RF}}$.

Under the assumption of perfect CSI, the HDA architecture does increase the MU-MIMO performance compared to a fully-digital MU-MIMO system with $M_{\mathrm{RF}}$ antenna elements. The analog beamforming forms a diagonally dominant $\tilde{\mathbf{H}}$ for the digital system. Hence, the condition number of $\tilde{\mathbf{H}}$ is smaller, and as a result, the MU-MIMO performance is better (see Section 2.1), compared to a channel given by a system with $M_{\mathrm{RF}}$ antenna elements directly sounding the physical channel. Naturally, the performance of the HDA architecture is upper bounded by that of a fully-digital architecture [14]. The performance gap depends, among other things, on the characteristics of the channel. Especially for mm-wave channels (see [21], [53] for the characteristics of such channels), an HDA system can achieve the same performance as a fully-digital system [14]. On the other hand, the HDA architecture has a much lower hardware complexity with respect to the fully-digital architecture. Also, the training overhead to estimate the necessary CSI is much less for the HDA architecture. These two advantages are discussed in the next section.

## 2.4. Application Scenarios of the HDA Architecture

As we have described the main idea of massive MIMO, we have seen the achievable performance gain and the possible disadvantages or costs. The two considered architectures,

the fully-digital architecture and the HDA architecture, have different advantages and disadvantages. In the following, we want to discuss the application scenarios for which the HDA architecture is recommended.

The hardware complexity of massive MIMO systems increases significantly as the number of antennas grows large. For some applications, complexity is a major factor and can make massive MIMO infeasible. Millimeter-wave systems are one important application scenario where the hardware complexity can be a limitation [14]. The high carrier frequency and the large channel bandwidth (equivalent to the sampling rate) make the hardware design more challenging compared to sub-6 GHz systems. Especially, the components of the RF chains are demanding. The cost and the power consumption of the ADCs/DACs and other RF parts are higher. If the increased cost and power consumption are scaled up by a large number of RF chains, a system can become infeasible. The HDA architecture is a solution to the hardware complexity increase. Its reduced number of RF chains makes the increase per chain tolerable. Additionally, the achievable MU-MIMO performance in mm-wave channels of hybrid systems is comparable to fully-digital systems [14]. In consequence, application scenarios where the hardware complexity is an issue and the performance gap is small are best suited for the HDA architecture. This can also be observed in the literature on mm-wave communication systems, where almost exclusively hybrid systems are investigated [10], [15], [17]–[21], [53]. A complete analysis of the complexity of the HDA architecture, including a comparison with the fully-digital architecture, follows in Chapter 3.

Besides the hardware complexity, the amount of training overhead is an important criterion of massive MIMO systems. In the previous section, we introduced JSDM as a solution to the training overhead issue in FDD massive MIMO systems. One can easily see, that the structure of the HDA architecture can be used for JSDM processing. The mathematical description of the precoding of JSDM (2.6) and the hybrid architecture (2.8) is very similar. The analog beamforming network in the HDA structure can be directly used for the beamforming stage of JSDM, $\mathbf{U}_{\mathrm{RF}} = \mathbf{B}$. The reduced dimension of the second stage MU-MIMO precoding of JSDM is equivalent to the reduced number of RF chains in the HDA architecture. The MU-MIMO precoder of JSDM can be implemented in the digital part of the HDA architecture, $\mathbf{W}_{\mathrm{BB}} = \mathbf{P}$. As mentioned in the introduction, the HDA architecture requires a special form of channel estimation which we call beam alignment (BA). Most BA algorithms proposed in the literature actually estimate the second-order statistics of the channel [28], [30], [31]. Thus, the BA is equal to the estimation of the covariance matrices in JSDM. The estimation of the effective channel can be done with standard channel estimation techniques similar to other MU-MIMO systems, e.g., using feedback. Since JSDM processing and the HDA architecture are so similar, the HDA

architecture can be used for application scenarios where the amount of training data is an issue, e.g., FDD networks.

An additional application of the HDA architecture is to extend existing non-massive MU-MIMO systems without the need to change other parts of those systems, e.g., the communication protocol. The analog beamforming network and a massive antenna array can be set up in front of the RF chains of an existing system. The HDA extension can boost the MU-MIMO performance of the system. The motivation behind such an implementation, in contrast to redesigning as a fully-digital massive MIMO system, is not a technological or fundamental limit but a limitation made by choice. One example of such a case is the limitation by a network standard. Standardization can require multiple years and can have an impact on millions of devices. A standard might not support the necessary protocol parts for the fully-digital massive MIMO processing, e.g., by limiting the maximum number of spatial streams. Nonetheless, it can still be possible to use an HDA extension without any change to the protocol. Wi-Fi is an example of such a protocol as can be seen in Chapter 6. The IEEE 802.11 standard might never support fully-digital massive MIMO processing because most of the Wi-Fi networks do not require such a large MU-MIMO gain, e.g., private WLANs. However, some Wi-Fi networks would benefit from the large MU-MIMO gain, e.g., business networks in large rooms with many users. An HDA extension is a solution to this problem.

In conclusion, the following list summarizes the application scenarios for the HDA architecture.

- *Hardware complexity limited systems:* for example, mm-wave 5G systems, systems where cost and power consumption is a significant factor.
- *CSI acquisition limited systems:* for example, FDD-based systems.
- *By design or protocol limited systems:* for example, current Wi-Fi systems.

# 3. Implementation of the Hybrid Digital-Analog Architecture

The implementation of a complete HDA communication system is complex and involves many tasks. However, many of them are similar to other massive MIMO systems or, in general, communication systems. The two parts which are specific to the HDA architecture are the analog beamforming network and parts of the signal processing. The design of the analog network involves the structure of the network and the components within the network. The signal processing includes algorithms both for the CSI acquisition phase and the communication phase (e.g., the precoding). The signal processing also calculates the beamforming weights which are applied to the controllable analog network. This chapter focuses on the HDA-specific parts of a system and their properties and is organized as follows. Section 3.1 introduces the two main candidates for the structure of the analog network. The structure is part of the hardware design but also has an impact on the signal processing implementation. In Section 3.2, we introduce and investigate signal processing algorithms for the hybrid architecture. The selected algorithms were proposed in the literature by other authors. They are summarized and it is explained, why these specific algorithms were chosen. The hardware of the analog network is investigated in Section 3.3. We consider different properties of the hardware and how they influence the system characteristics. The last section, Section 3.4, gives an overview of some aspects of digital hardware. The chapter is based on the publications [35], [36], [25], and [27].

In this chapter, we use different constraints for the radiated output power. We differentiate between three power constraints: a total radiated power constraint, a per RF chain power constraint, and a per antenna element power constraint. The total radiated power constraint limits the sum power transmitted by all antenna elements to a given value $P_{\text{tot}}$. The power transmitted by each antenna element or RF chain can vary as long as the sum is equal to $P_{\text{tot}}$. In an extreme case, the total output power could also be radiated from a single element. Such a constraint could be dictated by a communication standard or a legal regulation. The per RF chain power constraint sets a limit on the maximum output power of each RF chain. The maximum total radiated power of a system under such a constraint is $P_{\text{tot,max}} = M_{\text{RF}} \cdot P_{\text{RF,max}}$. If the transmission signal processing results in different

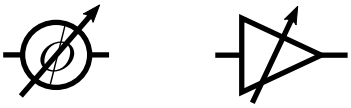
amplitudes per RF chain, the largest signal amplitude would be scaled to $P_{\mathrm{RF,max}}$ and the smaller ones would result in lower output power. In such a situation the total radiated output power $P_{\mathrm{tot}}$ would be less than $P_{\mathrm{tot,max}}$. The per antenna element power constraint is similar to the per RF chain constraint just with the maximum output power limited for each antenna element. The maximum total radiated power with a per antenna element power constraint is $P_{\mathrm{tot,max}} = M \cdot P_{\mathrm{elem,max}}$. The largest amplitude of the transmitted signal vector $\hat{\mathbf{x}}$ would be equivalent to the maximum power per element $P_{\mathrm{elem,max}}$. The total radiated power is always $P_{\mathrm{tot}} \leq P_{\mathrm{tot,max}}$. The per RF chain power constraint and the per antenna element power constraint could be dictated by the hardware of the RF system. For example, an amplifier in front of each antenna element could limit the maximum power per element.

The layout of the antenna array defines how the $M$ antenna elements are distributed within space. A uniform linear array (ULA) has antenna elements along a single dimension of a coordinate system. All elements sit on a line along this dimension with a defined and constant distance between the elements [54, Ch. 22]. Without loss of generality, we assume a Cartesian coordinate system with the axes $x$, $y$, and $z$. If the array is arranged along the $x$-axis, the antenna could steer beams along the azimuth angle of the spherical coordinate systems. The array factor is constant along the elevation angle. The beam width and other parameters, which will be used below, depend on $M$. If the antenna elements are not arranged along a single axis but along multiple dimensions (e.g., in a uniform rectangular array (URA)), this is not true anymore. In the following, we assume a ULA. The signal processing and also the structure of the analog network could be extended to more complex antenna layouts but this is not part of this thesis.

## 3.1. Structure and Components of the Analog Network

The analog part of the hybrid approach is realized by the analog network which connects a large number of antennas $M$ to a smaller number of RF chains $M_{\mathrm{RF}}$ and the digital signal processing part as shown in the block diagram in Figure 2.1. This dimensionality reduction $M$ to $M_{\mathrm{RF}}$ and its realization by the beamforming network are the core of the HDA concept. The analog network can mainly vary in two different ways: the structure of the interconnections within the network, and the type of the controllable component in every interconnection. Different structures and component types have been investigated in the literature [14], [17], [22]–[25]. This section introduces the components and the structures but does not analyze their characteristics. We focus on the two main structures and the most typical component type. The communication performance is presented in the signal processing section. The more hardware-related characteristics, e.g., the power efficiency,

Figure 3.1.: Block diagram symbol of a power combiner/divider.



(a) (b)

Figure 3.2.: Block diagram symbol of (a) a phase shifter and (b) an amplitude modulator.

are shown in the hardware section. For simplicity, we consider the transmission case (i.e., the downlink) in the mathematical modeling. The beamforming by the analog network is mathematically modeled as the matrix $\mathbf{U}_{\mathrm{RF}} \in \mathbb{C}^{M \times M_{\mathrm{RF}}}$. We denote the signals of the RF chains as $\mathbf{x}_{\mathrm{RF}} = [x_{\mathrm{RF},1}, \cdots, x_{\mathrm{RF},M_{\mathrm{RF}}}]$. Following, the beamformed output vector $\hat{\mathbf{x}}$ can be written as

$$\hat{\mathbf{x}} = \mathbf{U}_{\mathrm{RF}} \cdot \mathbf{x}_{\mathrm{RF}} \tag{3.1}$$

$$\hat{\mathbf{x}} = \sqrt{\alpha_{\mathrm{com}}} \cdot \tilde{\mathbf{U}}_{\mathrm{RF}} \cdot \sqrt{\alpha_{\mathrm{div}}} \cdot \mathbf{x}_{\mathrm{RF}} \tag{3.2}$$

where $\alpha_{\mathrm{com}}$ and $\alpha_{\mathrm{div}}$ represent the power combiners and dividers in the network, respectively. $\tilde{\mathbf{U}}_{\mathrm{RF}}$ is the plain beamforming matrix without the effect of the power combiners/dividers.

The analog network consists of components creating the tree structure and controllable components in every branch of the structure. Power combiners/dividers are the hardware components that span the tree structure of the network. A power combiner merges multiple signals into one and a power divider splits one signal into multiple. In this work, we use a common symbol, shown in Figure 3.1, for the combiner and the divider. $\alpha_{\mathrm{com}}$ and $\alpha_{\mathrm{div}}$ model the physical transmission factor of the hardware components. The most common hardware implementation is the so-called Wilkinson divider. It is reciprocal and not lossless [55, Ch. 7]. $\alpha_{\mathrm{com}}/\alpha_{\mathrm{div}}$ of a Wilkinson divider is equal to its port ration, e.g., $\alpha_{\mathrm{com}} = 1/2$ for a two to one combiner. Section 3.3 describes the power combiners/dividers in more detail.

The controllable components in the branches of the analog network realize the beamforming function. Each non-zero element of the beamforming matrix $\tilde{\mathbf{U}}_{\mathrm{RF}}$ represents a controllable component. The type of component can either be an on-off switch, a phase shifter, or a vector modulator. Each type results in a different restriction for the beamforming and the elements of the matrix $\tilde{\mathbf{U}}_{\mathrm{RF}}$. For the following, we assume that $\tilde{\mathbf{U}}_{\mathrm{RF}}$ is normalized. The elements of $\tilde{\mathbf{U}}_{\mathrm{RF}}$ are limited by an on-off switch to be $[\tilde{\mathbf{U}}_{\mathrm{RF}}]_{m,n} \in \{0, 1\}$, by a phase shifter to be $[\tilde{\mathbf{U}}_{\mathrm{RF}}]_{m,n} = e^{j\phi_{m,n}}$, and by a vector modulator to be $[\tilde{\mathbf{U}}_{\mathrm{RF}}]_{m,n} = a_{m,n}e^{j\phi_{m,n}}$.

Here $\phi_{m,n} \in [-\pi, \pi]$, and $a_{m,n} \in [0, 1]$ are the selected phase and amplitude. A vector modulator is often implemented as a combination of a phase shifter and an amplitude modulator. Figure 3.2 shows the block diagram symbols of a phase shifter and an amplitude modulator. The values of the elements of $\tilde{\mathbf{U}}_{\mathrm{RF}}$ and, hence, the settings for the controllable components are determined by the signal processing algorithm.

This thesis focuses on vector modulators as controllable components. Vector modulators enable the highest degree of freedom in the algorithms but are more complex to implement than the other options. The authors in [23] have extensively studied the performance/complexity/power consumption trade-offs of the use of on-off switches and phase shifters in the analog network. According to their work, systems based on switches are less complex but achieve a lower spectral efficiency compared to systems based on phase shifters. The power consumption in relation to spectral efficiency is the same for both types. The results of the performance evaluation also apply to vector modulators since they are commonly phase shifters extended by amplitude modulators. A vector modulator with an amplitude setting of 1 is equivalent to a phase shifter. The higher complexity due to the added amplitude modulator is acceptable with state-of-the-art hardware. The hardware platforms which are the basis for the system performance investigations in Chapters 5 and 6 both support amplitude and phase control. Also, the industry offers multiple beamforming integrated circuits (ICs) with phase and amplitude control for the mm-wave application scenario.[1,2,3,4,5] Thus, the focus on vector modulators is plausible.

The most important design aspect of the analog network is its structure. The two main network structures are the fully-connected (FC) and the one-stream-per-subarray (OSPS) structure. The FC structure is shown in Figure 3.3(a). Every RF chain is connected to every antenna element. Hence, $\tilde{\mathbf{U}}_{\mathrm{RF}}$ is a full matrix with the form

$$\tilde{\mathbf{U}}_{\mathrm{RF}}^{\mathrm{FC}} = [\tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_2, \cdots, \tilde{\mathbf{u}}_{M_{\mathrm{RF}}}], \tag{3.3}$$

[1] NXP MMW9004KC, 24 GHz to 27 GHz 4-channel beamforming IC, no publicly available data sheet, `https://www.nxp.com/products/radio-frequency/rf-power/rf-cellular-infrastructure/5g-mmwave/24-25-27-5-ghz-4-channel-analog-beamforming-integrated-circuit:MMW9004KC` Retrieved 12 November 2021

[2] Anokiwave AWMF-0165, 24 GHz to 27 GHz 8-channel beamforming IC, no publicly available data sheet, `https://anokiwave.com/products/awmf-0165/index.html` Retrieved 12 November 2021

[3] Renesas F5268, 24 GHz to 27 GHz 8-channel beamforming IC, only short-form data sheet publicly available, `https://www.renesas.com/eu/en/products/rf-products/phased-array-beamformers/f5268-8-channel-dual-polarization-trx-beamformer-ic-2425ghz-275ghz` Retrieved 12 November 2021

[4] MixComm SUMMIT2629, 26 GHz to 29 GHz 8-channel beamforming IC, no publicly available data sheet, `https://mixcomm.com/mixcomm-products/` Retrieved 12 November 2021

[5] Analog Devices ADMV4821, 24 GHz to 29 GHz 16-channel beamforming IC, no publicly available data sheet, `https://www.analog.com/en/products/admv4821.html` Retrieved 12 November 2021

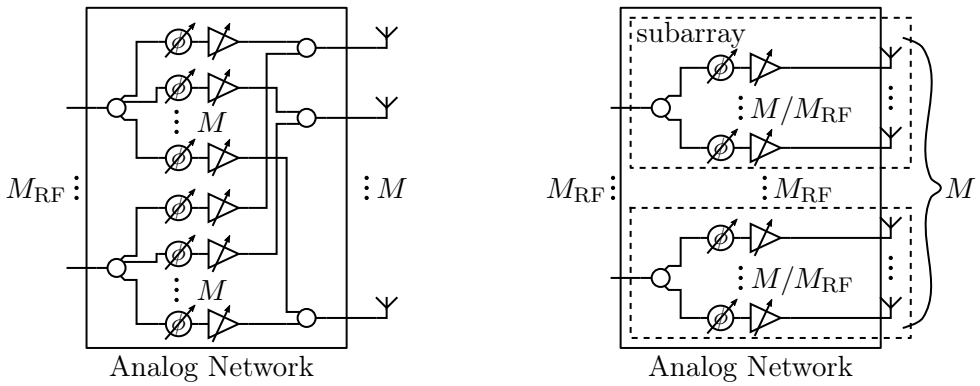



Figure 3.3.: Two possible structures for the analog network: (a) fully-connected (FC) structure, (b) one-stream-per-subarray (OSPS) structure.

where the vector $\tilde{\mathbf{u}}_k \in \mathbb{C}^M$ is the beamforming vector for the $k$-th RF chain. In the FC case, the power combiner/divider transmission factor results in $\alpha_{\mathrm{com}} = 1/M_{\mathrm{RF}}$ and $\alpha_{\mathrm{div}} = 1/M$. Figure 3.3(b) shows the OSPS structure. In this structure, one RF chain is only connected to a subset of the antenna elements and every element is only connected to one RF chain. This yields to $\tilde{\mathbf{U}}_{\mathrm{RF}}$ having the form

$$\tilde{\mathbf{U}}_{\mathrm{RF}}^{\mathrm{OSPS}} = \begin{bmatrix} \tilde{\mathbf{u}}_1 & 0 & ... & 0 \\ 0 & \tilde{\mathbf{u}}_2 & ... & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & ... & \tilde{\mathbf{u}}_{M_{\mathrm{RF}}} \end{bmatrix}, \tag{3.4}$$

where $\tilde{\mathbf{u}}_k \in \mathbb{C}^{M/M_{\mathrm{RF}}}$ is again the beamforming vector for the $k$-th RF chain but with the size $M/M_{\mathrm{RF}}$. In the OSPS case, $\alpha_{\mathrm{com}} = 1$ and $\alpha_{\mathrm{div}} = M_{\mathrm{RF}}/M$. In the OSPS structure, only $M/M_{\mathrm{RF}}$ antenna elements are connected to a single RF chain. In contrast, all $M$ elements are connected to each RF chain in the FC structure. Therefore, the antenna array gain for a single RF chain of the OSPS structure is lower than in the FC structure. The structure also influences the narrowest possible beam width. Although, we can not make a general remark on the two structures since the minimum beam width also depends on the array configuration (i.e., the arrangement of the elements in the array).

The FC structure and the OSPS structure can be seen as the two "extreme" cases of all possible structures. In the following sections, we will compare the two structures to gain insights into their advantages and disadvantages. Nonetheless, a hybrid system could use a

structure between the two extremes to optimize the system performance under the given constraints.

## 3.2. Signal Processing

This section will introduce the signal processing algorithms for the HDA architecture. The algorithms are realized by the digital processing and the analog network. We investigate algorithms for the two phases of communication in an HDA system. The first phase in a wireless communication process is the acquisition of users by the BS/AP. This phase is called initial access and includes the initial measurement of the CSI. For a hybrid system, we call the initial acquisition of the CSI beam alignment (BA). Section 3.2.1 describes the corresponding algorithms. The second phase is the data communication phase during which the BS/AP and the UEs exchange the user data. In a hybrid system, this includes the MU-MIMO precoding/combining. Section 3.2.2 will cover the precoding algorithms for the data communication in an HDA system.

### 3.2.1. Beam Alignment Algorithms

#### Motivation and Introduction

Compared to the fully-digital massive MIMO architecture, the hybrid architecture complicates the initial CSI acquisition. The CSI in a massive MIMO system can be very large due to the large number of antenna elements at the BS/AP and, potentially, at the UEs. A hybrid system, in contrast to a fully-digital system, can not measure the CSI instantaneously due to the low number of RF chains. A hybrid system can only measure per orthogonal training resource a low-dimensional representation of the large channel matrix.[6] Depending on the BA algorithm, the system uses multiple training resources to estimate the full channel or a certain representation of it. The overall resource usage (e.g., the training time) necessary for the CSI estimation can become very large and limit the sum rate of the system.

The mm-wave application scenario poses a second challenge for the initial access phase. The isotropic path loss (as seen by an omnidirectional antenna) increases with the square of the frequency [54, Section 16.6]. The isotropic path loss in the mm-wave frequency range can be very high and prohibit the direct omnidirectional measurement of the channel. Hence, a mm-wave system requires a directional antenna gain (i.e., beamforming) to achieve a sufficient SNR during the CSI measurement or the data communication.

[6] An orthogonal resource can be a time slot, a sequence, or a frequency block.

Almost all algorithms proposed in the literature [17], [28], [30], [31] solve these two issues by, first, applying some kind of beamforming during the measurements and, second, by estimating only a subset of the channel information. Commonly, they estimate the AoD/AoA pairs and the complex gains of the strongest paths for each user. Figuratively, they align the beam directions (AoD/AoA) between the BS/AP and the UEs which is why we call the process beam alignment. The justification for this approach is the so-called "sparsity" of the mm-wave channel. Extensive channel measurements have shown that mm-wave channels typically exhibit on average up to 3 multipath components, each corresponding to a scattering cluster with small delay/angle spreading [56]. As a result, a suitable BA algorithm only needs to identify a very small subset of the channel.

As mentioned during the introduction of JSDM in Section 2.2, also other application scenarios (e.g., FDD systems) pose problems for the massive MIMO CSI acquisition. The authors in [45] proposed with JSDM a similar split of the CSI in the second-order statistics and a low-dimensional instantaneous channel. The AoD/AoA pairs and the gains of all the strong paths in the channel are equivalent to the second-order statistics. Thus, the algorithms proposed for the mm-wave BA can be used for the estimation of the second-order statistics for JSDM. This is true as long as all the strongest paths can be estimated within the given training time.

We categorize BA algorithms into two types, basic and advanced algorithms. Basic schemes run a search-like process. Such a scheme is, for example, used in the IEEE 802.11ad standard [57] (60 GHz Wi-Fi). Usually, the training time of basic schemes does not scale well with the number of antenna elements. Advanced schemes use more sophisticated signal processing but scale better with the system size. In the following, we introduce one basic and one advanced algorithm.

Without loss of generality, in the following, we are considering training slots as the orthogonal training resource. The training resource could also be a frequency resource or any other orthogonal resource. For simplicity, we consider only UEs with a single RF chain.

#### Basic Algorithm

The most basic solution for the initial acquisition is an exhaustive search. The analog network of the hybrid system and the antenna array are used as a standard phased array with Fourier-matrix coefficients [54, Chapter 22]. As an example, the antenna gain of two of such beam patterns of a 16-element array is shown in Figure 3.4. The array forms narrow beams with a certain beam direction and a given beamwidth $\Delta\theta$. $\Delta\theta$ is dependent on the number of antenna elements in the beamforming dimension and the direction of the beam. The antenna directivity (equally the array factor, if we ignore the single element pattern)

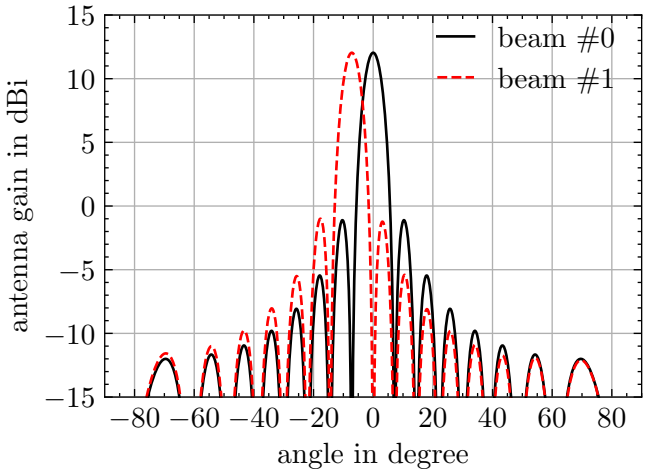



Figure 3.4.: Antenna array gains of two adjacent single beam pattern examples (Fourier-matrix coefficients) of a 16-element array ($M' = 16$).

generated with the phased array is equivalent to the maximum possible directivity using a set number of elements. A codebook $\mathcal{C}_{\mathrm{BS}}$ of beamforming vectors for the BS/AP is formed to cover the angular interval with a given number $M'$ of equidistant beams. A common choice for the number of equidistant beams, which is also the size of the codebook $\mathcal{C}_{\mathrm{BS}}$, is the number of elements in the beamforming direction. For a ULA, the FC analog network structure would have a $\mathcal{C}_{\mathrm{BS}}$ size of $M' = M$ and the OSPS structure a size of $M' = M/M_{\mathrm{RF}}$. UEs with multiple antenna elements would form equivalent codebooks $\mathcal{C}_{\mathrm{UE}}$ with the size $N_{\mathrm{UE}}$, where $N_{\mathrm{UE}}$ is the number of antenna elements in the beamforming dimension per UE. Since both the BS/AP and the UEs are equipped with antenna arrays, some form of synchronization is necessary. Assuming the existence of a side channel for synchronization, the exhaustive search algorithm checks every combination of codewords of $\mathcal{C}_{\mathrm{BS}}$ and $\mathcal{C}_{\mathrm{UE}}$ and stores the corresponding channel gain. The BS/AP can set up/measure one codeword of $\mathcal{C}_{\mathrm{BS}}$ per RF chain. If the search requires less than the coherence time of the channel it measures the instantaneous full CSI. The exhaustive search requires $\frac{M'}{M_{\mathrm{RF}}} \cdot N_{\mathrm{UE}}$ training slots per UE. If the training is done in the uplink, e.g., in a TDD system, the UEs can not be trained simultaneously and the training requires in total $\frac{M'}{M_{\mathrm{RF}}} \cdot N$ training slots. The downlink training of multiple UEs, e.g., in an FDD system, can be simultaneous. In this case, the number of training slots is $\frac{M'}{M_{\mathrm{RF}}} \cdot N_{\mathrm{UE}}$. The required resource usage of the feedback towards the BS/AP depends on the type of required CSI. If the instantaneous CSI is sent back to the BS/AP, the total amount of feedback would scale with $M' \cdot N$.[7]

[7]The total amount also depends on the resolution and the data type of the feedback.

If only the AoD information is required at the BS/AP, the total amount of feedback only scales with $N/N_{\mathrm{UE}}$, i.e., the number of users.

For a further analysis of the exhaustive search algorithm, see [58].

#### Advanced Algorithm

In contrast to the basic scheme described above, one goal of more advanced BA algorithms is to use a minimum of training resources. The literature considers mm-wave systems as the main application of the hybrid architecture. Thus, most of the works on BA algorithms are assuming mm-wave channels. Due to the sparse nature of mm-wave channels, compressed sensing is considered a powerful technique to reduce the number of training slots. There are different algorithms proposed in the literature, e.g., [28], [29], [31], [59]. This chapter presents the algorithm proposed in [31], [59]. The algorithm was not developed by the the author of this thesis. We use this algorithm in the systems which will be presented in Chapter 5 and Chapter 6 and describe it here as a reference. The simulation results given in this section are the results of a simulation implemented by the author of this thesis. The approach of [31] is for orthogonal frequency division multiplexing (OFDM) systems, whereas [59] uses pseudo-noise sequences in the time-domain as a training resource (e.g., for systems using single-carrier modulation). Both versions of the algorithm estimate the second-order statistics of the channel. The coherence time of the second-order statistics is much longer than that of the instantaneous CSI. The algorithm uses power measurements, instead of phase and amplitude measurements, for the estimation. Both features make the scheme robust to large Doppler spreads and fast variations of the instantaneous CSI.

The scheme works with measurements acquired both in the downlink and in the uplink. If the channel is measured in the downlink, the UEs estimate their AoA/AoD pair and send the information back to the BS/AP using a feedback channel. The amount of the feedback scales with the number of users $N/N_{\mathrm{UE}}$ and not with the number of antenna elements ($M$ or $N$). Due to the downlink measurements, all users can estimate their channel simultaneously. In a TDD system, the scheme could estimate the channel also in the uplink and use channel reciprocity to avoid feedback. Although, in the uplink, not all users can be trained simultaneously in the same training slot. The uplink training requires additional resources (time, frequency, or another orthogonal resource) to train all users. The authors of [31], [59] recommend the estimation using downlink measurements.

In the following, we describe the algorithm in the downlink case and for the FC structure. The analysis holds also for the OSPS structure but the number of elements $M$ would be replaced by $M/M_{\mathrm{RF}}$. The complete initial acquisition phase consists of two parts: a measurement/estimation period and a feedback period. This structure is compatible with

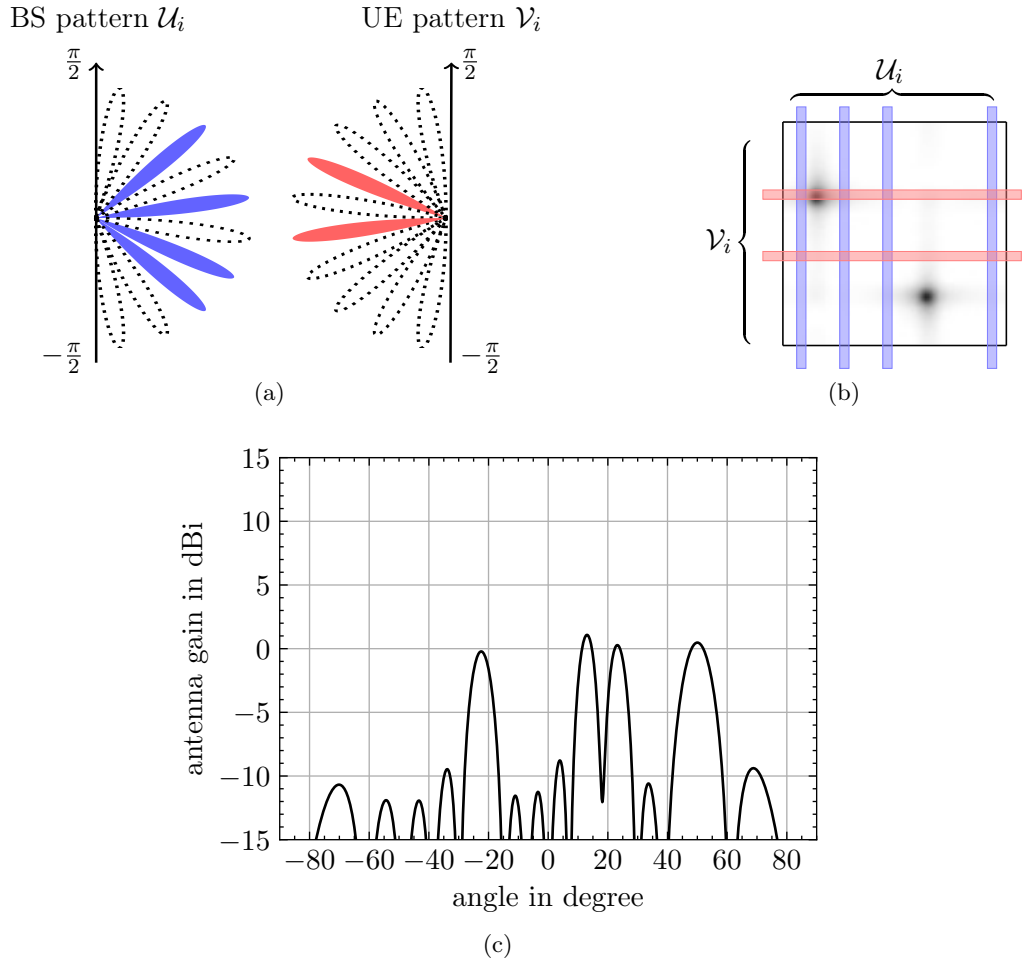



Figure 3.5.: Illustration of a BA training slot using 4 directions at the BS and 2 at the UE. (b) and (c) visualize the channel and the pattern subsets [31]. (c) is the antenna array gain of a 4-direction beam pattern of a 16-element array.

the 5G frame structure [60] and the IEEE 802.11ad beam training [57].[8] The BA process is done during the estimation period. The BS/AP broadcasts training signals during the training slots. The BS/AP uses the analog beamforming network to set up a pseudo-random beam pattern per training slot. The set of all random patterns at the BS/AP is the transmit beamforming codebook $\mathcal{C}_{\mathrm{T}}$). The codebook is known by the users. A UE measures during the training with its own codebook of receive beam patterns $\mathcal{C}_{\mathrm{R}}$. An illustration of one slot with selected codewords for the BS/AP $\mathcal{U}_i \in \mathcal{C}_{\mathrm{T}}$ and a UE $\mathcal{V}_i \in \mathcal{C}_{\mathrm{R}}$ is shown in Figures 3.5(a) and 3.5(b). The BS/AP uses different codewords/patterns per RF chain in each training slot. In case a UE is equipped with multiple RF chains, it also uses a unique codeword per RF chain. A codeword or single beam pattern consists of a pseudo-random selection of a given number of directions. The number of probed directions is $\kappa_u$ and $\kappa_v$ at the BS/AP and the UEs, respectively. The beamforming coefficients $\tilde{\mathbf{v}}_i$ and $\tilde{\mathbf{u}}_i$ of such patterns $\mathcal{V}_i$ and $\mathcal{U}_i$ can be written as

$$\tilde{\mathbf{u}}_i = \mathbf{F}_{\mathrm{BS}} \frac{\check{\mathbf{u}}_i}{\sqrt{\kappa_u}} \quad \text{and} \quad \tilde{\mathbf{v}}_i = \mathbf{F}_{\mathrm{UE}} \frac{\check{\mathbf{v}}_i}{\sqrt{\kappa_v}}, \tag{3.5}$$

where $\mathbf{F}_{\mathrm{BS}}$ and $\mathbf{F}_{\mathrm{UE}}$ are the discrete Fourier transform matrices at the BS/AP and the UEs, and where $\check{\mathbf{u}}_i \in \mathbb{C}^M$ and $\check{\mathbf{v}}_i \in \mathbb{C}^{N_{\mathrm{UE}}}$ are all-zero vectors with a 1 at components corresponding to the selected directions of $\mathcal{V}_i$ and $\mathcal{U}_i$, respectively. The elements of the discrete Fourier transform matrices $\mathbf{F}_{\mathrm{BS}}$ and $\mathbf{F}_{\mathrm{UE}}$ are given by

$$[\mathbf{F}_{\mathrm{BS}}]_{m,m'} = \frac{1}{\sqrt{M}} e^{j2\pi(m-1)(\frac{m'-1}{M}-\frac{1}{2})}, m, m' \in [M], \tag{3.6a}$$

$$[\mathbf{F}_{\mathrm{UE}}]_{n,n'} = \frac{1}{\sqrt{N_{\mathrm{UE}}}} e^{j2\pi(n-1)(\frac{n'-1}{N_{\mathrm{UE}}}-\frac{1}{2})}, n, n' \in [N_{\mathrm{UE}}]. \tag{3.6b}$$

The array gain for an exemplary beam pattern with 4 directions is shown in Figure 3.5(c). The number of probed directions is a trade-off between the beamforming gain of the patterns and the necessary training time. The smaller the number of directions is, the higher is the beamforming gain. Channels with a higher path loss, e.g., due to larger cell size, require a higher beamforming gain. In the extreme case of a single direction per pattern, the algorithm becomes equivalent to the exhaustive search algorithm. Besides the beam patterns using a pseudo-random selection of directions, the estimation can also be based on patterns created by a random phase per element.

A UE measures the received power per codeword/beamforming pattern (so for one training slot). In the frequency-domain version of the algorithm, the BS/AP uses orthogonal OFDM symbols as training signals. The training signals in the time-domain version are unique

[8] For 5G, the synchronization signal block (SSB) can be used for the estimation and the physical random access channel (PRACH) for the feedback.

pseudo-noise sequences. Each RF chain/pattern codeword pair of the BS/AP is assigned with a unique training sequence. A UE measures the received power by using matched filters for the pseudo-noise sequences or the training OFDM symbols. Over $T_{\mathrm{BA}}$ training slots a UE obtains a total number of $M_{\mathrm{RF}} \cdot T_{\mathrm{BA}}$ equations, which can be written in the form

$$\mathbf{q}_k = \mathbf{B}_k \cdot \mathrm{vec}(\mathbf{\Gamma}_k) + \zeta(P_{\mathrm{tot}}) \cdot \mathbf{1} + \mathbf{w}_k, \tag{3.7}$$

where $\mathbf{q}_k \in \mathbb{R}^{T_{\mathrm{BA}}}$ consists of all the $T_{\mathrm{BA}}$ statistical power measurements, $\mathbf{B}_k \in \mathbb{R}^{T_{\mathrm{BA}} \times M' N_{\mathrm{UE}}}$ is uniquely defined by the pseudo-random beamforming codebook of the BS/AP and the local beamforming codebook of the $k$-th UE, $\mathbf{\Gamma}_k$ is the beam-domain covariance matrix of the channel of the $k$-th user, $\zeta(P_{\mathrm{tot}})$ denotes a constant whose value is a function of the total radiated power, and $\mathbf{w}_k \in \mathbb{R}^{T_{\mathrm{BA}}}$ denotes the residual measurement fluctuations.

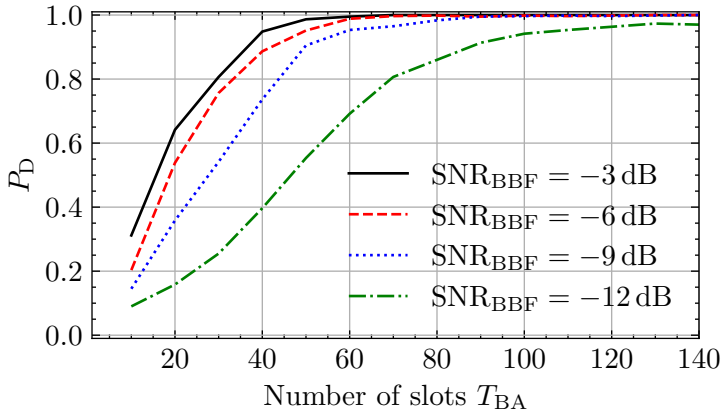
As discussed in [31], [59], the elements of $\mathbf{\Gamma}_k$ are always non-negative and a simple least-squares problem

$$\mathbf{\Gamma}_k^{\star} = \underset{\mathbf{\Gamma}_k \in \mathbb{R}_+^{N_{\mathrm{UE}}}}{\arg\min} \left\| \mathbf{B}_k \cdot \mathrm{vec}(\mathbf{\Gamma}_k) + \zeta(P_{\mathrm{tot}}) \cdot \mathbf{1} - \mathbf{q}_k \right\|^2 \tag{3.8}$$

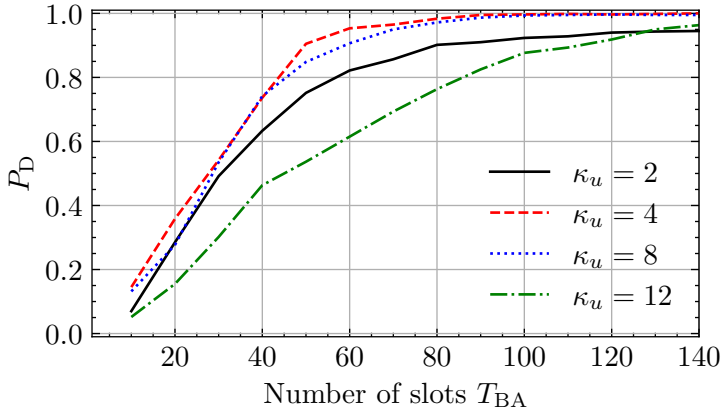
is sufficient to recover the solution $\mathbf{\Gamma}_k^{\star}$. The optimization problem of equation (3.8) is generally called *Non-Negative-Least-Squares* and it has well-investigated numerical solutions. We assume a successful BA process if the largest component in $\mathbf{\Gamma}_k^{\star}$ coincides with the actual strongest path of the $k$-th UE. As $\mathbf{\Gamma}_k$ is the covariance matrix of the channel in the beam-domain, the indices of the largest component are equivalent to the AoA and AoD. $\mathbf{\Gamma}_k$ is given in the beam-domain and has a size of $M \times N_{\mathrm{ue}}$ (see [59] for the exact definition). Following, the grid on which the AoD and AoA are estimated has $M$ and $N$ possible points defined by the Fourier transformation matrices $\mathbf{F}_{\mathrm{BS}}$ and $\mathbf{F}_{\mathrm{UE}}$. For example, the angle of the AoD at the BS/AP is given by $\phi_{\mathrm{AoD,k}} = \arcsin(2 \cdot j/M - 1)$ where $j$ is the index of the row of the maximum of $\Gamma_k$.

The number of measurements $T_{\mathrm{BA}}$ required to successfully estimate the AoA and AoD depends on the number of antenna elements $M$ and $N_{\mathrm{UE}}$, the number of probed directions $\kappa_u$ and $\kappa_v$, the number of RF chains $M_{\mathrm{RF}}$, and on the measurement SNR. $T_{\mathrm{BA}}$ may differ from user to user, depending on the individual SNR and on the number of receiver RF chains (assumed to be one in this section).

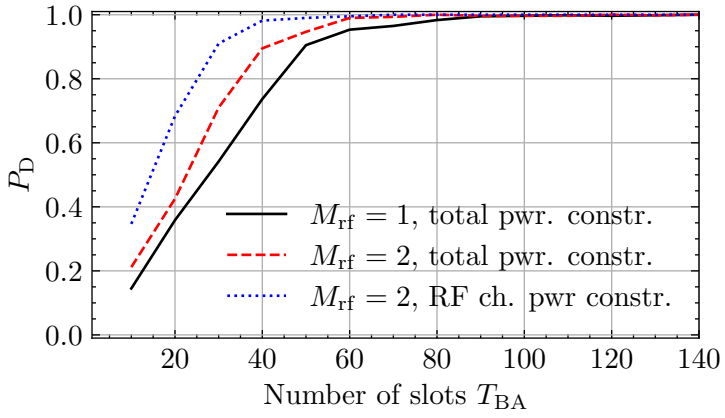
We have run numerical simulations of the time-domain version of the described algorithm to evaluate the performance. The simulation is based on the theoretical channel model as defined in [59]. The channel consists of three multipath components (equivalent to scattering clusters) with the relative strengths 1, 0.6, and 0.4. The component coefficients are modeled as Rice fading. The strength ratios between the line-of-sight (LOS) and non-




(a)




(b)




(c)

Figure 3.6.: Detection probability $P_D$ of the advanced BA scheme with $M = 16$, $N = 8$, and $\kappa_v = 4$. (a) for different $\mathrm{SNR}_{\mathrm{BBF}}$ values with $\kappa_u = 4$ and $M_{\mathrm{RF}} = 1$. (b) for different $\kappa_u$ values with $\mathrm{SNR}_{\mathrm{BBF}} = -9\,\mathrm{dB}$ and $M_{\mathrm{RF}} = 1$. (c) for different $M_{\mathrm{RF}}$ values and power constraints with $\mathrm{SNR}_{\mathrm{BBF}} = -9\,\mathrm{dB}$ and $\kappa_u = 4$.

line-of-sight (NLOS) components are 100, 10, and 0 for the three multipath components, respectively.[9] The simulation also includes a Doppler shift for all paths which is equivalent to a speed of $10\,\mathrm{m\,s^{-1}}$, a signal bandwidth of 0.8 GHz, and a carrier frequency of 40 GHz. To investigate the influence of the channel quality/strength on the performance, we use the SNR before beamforming ($\mathrm{SNR_{BBF}}$) as a metric for the channel. We define the $\mathrm{SNR_{BBF}}$ as

$$\mathrm{SNR_{BBF}} = \frac{P_{\mathrm{tot}} \sum_{l=1}^{L} \gamma_l}{N_0 \cdot B}, \tag{3.9}$$

where $P_{\mathrm{tot}}$ is the total radiated power of the transmitter, $L$ is the number of multipath components, $\gamma_l$ denotes the strength of the $l$-th component, $N_0$ is the power spectral density of the noise at the receiver, and $B$ is the total used bandwidth. The $\mathrm{SNR_{BBF}}$ is the communication data SNR obtained when a single data stream is transmitted through a single BS/AP antenna element and is received at a single UE antenna element (isotropic transmission) over the whole bandwidth. The simulated hybrid system has $M = 16$ antenna elements and, if not stated otherwise, $M_{\mathrm{RF}} = 1$ RF chain. The UE has $N_{\mathrm{UE}} = 8$ antenna elements and one RF chain. The beamforming codebook at user side $\mathcal{C}_{\mathrm{R}}$ probes $\kappa_v = 4$ directions per codeword. The pseudo-noise sequence has a length of 64 samples and one sequence per training slot is used.

Figure 3.6 shows the detection probability $P_D$ (i.e., the probability of successfully estimating the AoA/AoD pair) over the number of training slots $T_{\mathrm{BA}}$. The performance depends on the $\mathrm{SNR_{BBF}}$ as can be seen in Figure 3.6(a). For the given system, the algorithm successfully ($P_D \geq 0.95$) estimates the strongest path with an $\mathrm{SNR_{BBF}} \geq -9\,\mathrm{dB}$ after approximately 60 slots. This is less than half of the slots which an exhaustive search algorithm would require (which is at least $M \cdot N_{\mathrm{UE}} = 128$ slots). The number of required training slots also depends on the number of probed directions in a single beamforming pattern $\kappa_u$. A higher $\kappa_u$ decreases the training time but also decreases the beamforming gain and, hence, the measurement SNR. One can see in Figure 3.6(b) that $\kappa_u = 4$ is optimal for an $\mathrm{SNR_{BBF}}$ of $-9\,\mathrm{dB}$. Both a higher and a lower $\kappa_u$ decreases the performance. Figure 3.6(c) shows the influence of the number of RF chains on the performance. For simplicity, we only consider the fully-connected structure of the analog network. The number of RF chains is strongly related to the chosen power constraint. With a total power constraint, the total output power is "shared" between the RF chains. The output power per RF chain is the total power $P_{\mathrm{tot}}$ divided by $M_{\mathrm{RF}}$. In contrast, a power constraint per RF chain increases the total output power of the system if $M_{\mathrm{RF}}$ increases. In Figure 3.6(c), one can see two effects. First, the increase from $M_{\mathrm{RF}} = 1$ to $M_{\mathrm{RF}} = 2$ under the per RF

[9] This is equivalent to one LOS path and two NLOS paths (one slightly stronger than the other).

chain power constraint reduces the required number of slots by half. The measurement SNR for each RF chain is independent of the total number of RF chains under the per RF chain power constraint. Hence, it is clear that the "quality" of each measurement does not change and that doubling $M_{\mathrm{RF}}$ doubles the number of measurements per slot. This of course reduces the number of slots by half. Second, with the total power constraint, when increasing $M_{\mathrm{RF}}$ the number of slots is still reduced but not by the same factor. The measurement SNR decreases due to the total power constraint when $M_{\mathrm{RF}}$ increases. A smaller SNR increases the required number of measurements for a successful BA, as can be seen in Figure 3.6(a). The higher number of RF chains still increases the number of measurements by $M_{\mathrm{RF}}$ but, due to the smaller SNR, the total number of measurements only decreases by a smaller factor.

After the UE has estimated the AoA/AoD pair of its strongest path, it feeds the AoD information back to the BS/AP. During the feedback period, the BS/AP is in a listening mode such that each UE sends a beamformed packet to the BS/AP. This packet contains basic information such as the UE ID and the beam indices of the selected AoD. From this moment on, the BS/AP and the UE are connected in the sense that, if the procedure is successful, they have achieved BA. In other words, they can communicate by aligning their beams along a multipath component with the estimated AoA/AoD and strong channel coefficient due to the beamforming gain of both arrays.

More details on the mathematical description of the algorithm, on the different parameters, and the performance can be found in [31], [59].

#### BA Algorithm Conclusion

The exhaustive search algorithm is easy to implement but the number of required training slots does scale with the product of $M'$ times $N_{\mathrm{UE}}$. This makes the algorithm infeasible especially for networks where both the BS/AP and the UEs have many antenna elements. Nonetheless, it might be an option for systems with smaller antenna arrays (e.g., $M < 16$ and $N < 4$). The advanced algorithm does reduce the number of required training slots significantly and makes it more suitable for systems with large antenna arrays. In the following, the advantages of the described algorithm are listed.

- It has a very high performance with respect to the required number of training slots. Also, when compared with the state-of-the-art [31], [59].

- It can be used in time-domain and frequency-domain systems and also in the uplink and the downlink.

- It is flexible with respect to the expected SNR regime. The beamforming patterns can be adjusted, e.g., $\kappa_u$ can be optimized to get the optimal $T_{\mathrm{BA}}$ for a given SNR.
- It is based on power measurements and, hence, robust to large Doppler spreads.

### 3.2.2. Data Communication Algorithms

After the initial acquisition of users for the BS/AP during the BA process, the system can switch to a communication phase. In this communication phase in the downlink, the user data is transmitted with MU-MIMO precoding from the BS/AP to the UEs. In the uplink, multiple UEs can transmit at the same time by using MU-MIMO combining at the BS/AP. In the following, we consider the downlink case and introduce an MU-MIMO precoding algorithm. Many precoding algorithms have been proposed in the literature, e.g., [18], [19], [27], [32], [33]. Some algorithms [32], [33] require perfect instantaneous CSI at the transmitter which, as described above, is usually not given in a hybrid system. Below, we introduce the hybrid precoding algorithm described in [27]. The author of this thesis was a co-author of this publication, although the algorithm was developed by the main author. The author of this thesis implemented the simulations presented in this section. The algorithm requires only the knowledge of the second-order statistics of the channel matrix $H$ which can be obtained by a BA algorithm as described above. The algorithm presented in [18] is similar but not as thoroughly investigated (e.g., in a co-simulation with the BA). For a comparison with [19], please see [27].

We assume that the BS/AP simultaneously schedules $K = M_{\mathrm{RF}}$ users which are selected by a simple directional scheduler [19]. The simple scheduler selects $K$ users which have similar power profiles and whose strongest AoDs are at least $\Delta\theta_{\min}$ away from each other. This minimum separation angle ensures that a single beam points only towards a single user. Consequently, the multi-user beamforming scheme at the BS/AP allocates equal power across these users. An investigation of the scheduler is not part of this thesis. For further information and an exemplary scheduler please refer to [19]. Through the hybrid precoding, the BS/AP can apply multi-user interference cancellation for the scheduled user set. The goal of the precoding algorithm is to maximize the MU-MIMO gain and to achieve the interference-free capacity (2.5). The algorithm optimizes the precoding matrix $W$ as introduced in (2.2). Due to the HDA architecture, as introduced in Section 2.3, the complete hybrid precoding $\mathbf{W}$ (2.8) consists of the analog beamforming matrix $\mathbf{U}_{\mathrm{RF}}$ and the digital baseband precoding matrix $\mathbf{W}_{\mathrm{BB}}$, with $\mathbf{W} = \mathbf{U}_{\mathrm{RF}} \cdot \mathbf{W}_{\mathrm{BB}}$.

The basic concept of the proposed algorithm is, that the analog beamforming of the BS/AP and the UEs point beams towards the strongest path defined by the AoD and AoA. In other words, they align their beams along the strongest path. Afterward, the digital

part of the signal processing tries to maximize the sum rate by applying additional digital precoding. To attain the beamforming coefficients the $k$-th UE points a standard phased array beam with Fourier-matrix coefficients towards the estimated strongest direction. The beamforming coefficients are given similar to (3.5) by

$$\tilde{\mathbf{v}}_k = \mathbf{F}_{\mathrm{UE}} \check{\mathbf{v}}_k, \tag{3.10}$$

where $\check{\mathbf{v}}_k \in \mathbb{C}^{N_{\mathrm{UE}}}$ is an all-zero vector with a 1 at the component corresponding to the AoA of the strongest path. Assuming enough users in the cell and the described scheduler, the BS/AP beamforming coefficients for the $k$-th UE along the strongest AoD with respect to the chosen AoA are

$$\tilde{\mathbf{u}}_k = \mathbf{F}_{\mathrm{BS}} \check{\mathbf{u}}_k, \tag{3.11}$$

where $\check{\mathbf{u}}_k \in \mathbb{C}^M$ is an all-zero vector with a 1 at the component corresponding to the strongest AoD direction of the $k$-th UE. The full beamforming matrix $\mathbf{U}_{\mathrm{RF}}$ at the BS/AP is formed by all $\tilde{\mathbf{u}}_k$ for $k \in [K]$ depending on the structure of the analog network as described in Section 3.1 (3.3) and (3.3).

The composite receive beamforming matrix $\tilde{\mathbf{V}} \in \mathbb{C}^{N \times K}$ of all $K$ UEs is given by

$$\tilde{\mathbf{V}} = \mathrm{diag}(\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \dots \tilde{\mathbf{v}}_K). \tag{3.12}$$

The total composite receive beamforming matrix is $\mathbf{V} = \sqrt{\alpha_{\mathrm{com}}^{\mathrm{UE}}} \tilde{\mathbf{V}}$ including the effect of the power combiners in the beamforming network at the receiver.

As written in Section 2.3, the analog beamforming reduces the channel dimension seen by the digital processing. The effective channel $\tilde{\mathbf{H}}$ for the digital precoder is

$$\tilde{\mathbf{H}} := \mathbf{V}^{\mathsf{H}} \cdot \mathbf{H} \cdot \mathbf{U}_{\mathrm{RF}} \in \mathbb{C}^{K \times K}. \tag{3.13}$$

In contrast to (2.7), here we consider also UEs with multiple antenna elements and an analog beamforming network. The instantaneous effective channel can be estimated by the BS/AP using channel reciprocity and standard channel estimation methods at the cost of ($K \cdot K \ll M \cdot N$) orthogonal resources (similar to the JSDM scheme [45] introduced in Section 2.2).[10] We propose two variants for the digital part of the hybrid precoding algorithm [27].

*Beam steering (BST) scheme*: the BST scheme is the simplest possible approach for the digital signal processing. The BS/AP ignores the multi-user interference after the beamforming. It directly transmits the data stream for the $k$-th UE using the $k$-th RF

[10] For example, one standard channel estimation scheme in an OFDM system is to use uplink pilot symbols.

chain. Therefore, the digital precoding matrix is given by $\mathbf{W}_{\mathrm{BB}}^{\mathrm{BST}} = \mathbf{I}_K$. In this case, an additional uplink channel estimation of $\tilde{\mathbf{H}}$ can be omitted. The eventual $M \times K$ BST precoder in (2.8) reads

$$\mathbf{W}^{\mathrm{BST}} = \mathbf{U}_{\mathrm{RF}} \cdot \mathbf{W}_{\mathrm{BB}}^{\mathrm{BST}} = \mathbf{U}_{\mathrm{RF}}. \tag{3.14}$$

*Baseband zero-forcing (BZF) scheme*: in the BZF scheme, we consider zero-forcing precoding for potential multi-user interference cancellation. To calculate the ZF precoding matrix $\mathbf{W}_{\mathrm{BB}}^{\mathrm{ZF}}$ the base station needs to estimate the lower-dimensional effective channel $\tilde{\mathbf{H}}$. As a result, the baseband precoding matrix $\mathbf{W}_{\mathrm{BB}}^{\mathrm{ZF}}$ can be written as

$$\mathbf{W}_{\mathrm{BB}}^{\mathrm{ZF}} = \tilde{\mathbf{H}}^{\mathsf{H}} \cdot \left(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^{\mathsf{H}}\right)^{-1} \cdot \Delta^{\mathrm{ZF}}, \tag{3.15}$$

where $\Delta^{\mathrm{ZF}} \in \mathbb{R}_{+}^{K \times K}$ is a diagonal matrix, taking into account the power constraint. The eventual BZF precoder is then given by

$$\mathbf{W}^{\mathrm{ZF}} = \mathbf{U}_{\mathrm{RF}} \cdot \mathbf{W}_{\mathrm{BB}}^{\mathrm{ZF}}. \tag{3.16}$$

For a more extensive mathematical analysis of the data communication algorithm, please refer to [27].

### Numerical Simulation

We have run numerical simulations of the described algorithm to evaluate the performance and to compare the two schemes for the digital precoding. We calculate the achievable asymptotic ergodic spectral efficiency [61] as the metric for the evaluation of the sum rate performance. We assume no inter-carrier interference and treat interference by other users as noise. Following, the per-user spectral efficiency of the $k$-th UE is given by

$$R_k = \mathbb{E}\left[\sum_{\omega} \log_2\left(1 + \frac{P_{k,\omega}|\mathbf{v}_k^{\mathsf{H}}\mathbf{H}_{k,\omega}(t)\mathbf{u}_{k,\omega}|^2}{|\sum\limits_{k' \neq k} \sqrt{P_{k',\omega}}\mathbf{v}_k^{\mathsf{H}}\mathbf{H}_{k',\omega}(t)\mathbf{w}_{k',\omega}|^2 + P_{\mathrm{N}}}\right)\right]. \tag{3.17}$$

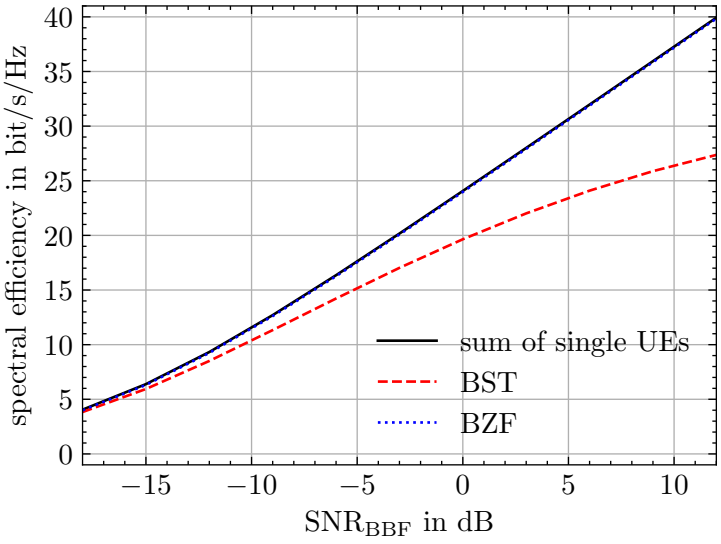
Where $\omega \in [F]$ is the subcarrier (SC) index and $F$ is the total number of SCs. In the simulation, we use $F = 256$ SCs. $\mathbf{H}_{k,\omega}(t)$ is the time-dependent channel, $\mathbf{v}_k$ is the receiver beamforming, $\mathbf{w}_{k,\omega}$ is the transmitter beamforming and precoding, and $P_{k,\omega}$ is the maximum output power of the transmitter for SC $\omega$ and user $k$. $P_{\mathrm{N}}$ is the noise power for the given SC bandwidth at a given device temperature including the noise figure of the receiver. We assign equal power to all SCs. The achievable asymptotic ergodic sum spectral efficiency

for the $K$ scheduled UEs is $R_{\text{sum}} = \sum_{k=1}^{K} R_k$. This metric does not include the overhead due to the BA and the instantaneous channel estimation, which is required by the BZF scheme. It is an upper bound. The sum spectral efficiency of a real system with a certain modulation and channel coding scheme will be lower. We analyze $R_{\text{sum}}$ over the $\text{SNR}_{\text{BBF}}$ as defined in (3.9).

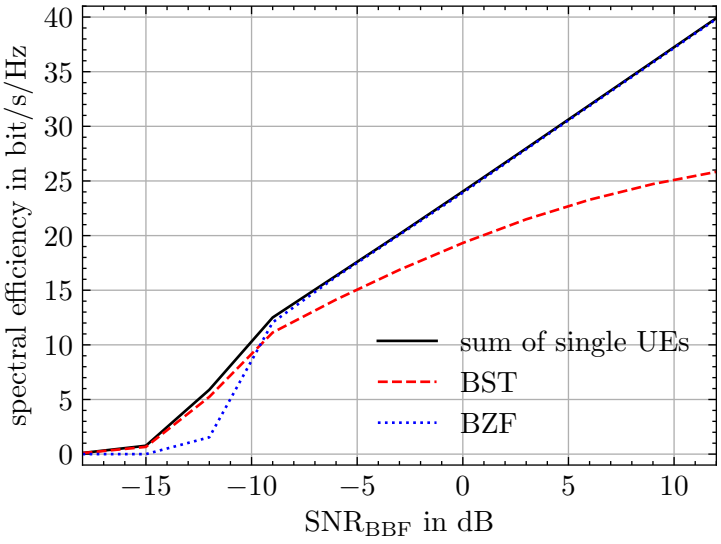
Similar to Section 3.2.1, the simulation is based on the theoretical channel model as defined in [59]. The channel consists of three multipath components with the relative strengths 1, 0.6, and 0.4. Since we consider the asymptotic ergodic spectral efficiency and coherent communication[11], the simulation does not include Rice fading and a Doppler shift. The SC bandwidth is 240 kHz and $F = 256$. We assume perfect CSI of the effective channel for the BZF scheme. The simulated hybrid system has $M = 64$ antenna elements and $K = M_{\text{RF}} = 4$ RF chains with am FC analog network. The UE has $N_{\text{UE}} = 8$ antenna elements and one RF chain. The system operates under a total radiated power constraint. The following graphs include the data of 120 users sets of each 4 users with random AoDs but an angular separation of $\Delta\theta_{\min} = 5°$.

Figure 3.7 shows the asymptotic ergodic sum spectral efficiency for both baseband schemes over the $\text{SNR}_{\text{BBF}}$. We compare the two schemes with the sum of the spectral efficiencies of the single users without any interference. The sum of the single users is the upper bound which can be achieved in such a system. In Figure 3.7(a), the transmitter has the perfect AoA/AoD information (i.e. the correct angles without an additional BA simulation). The BZF scheme can achieve over the whole $\text{SNR}_{\text{BBF}}$ range a sum spectral efficiency very close to the optimal sum of the single users. With the perfect AoA/AoD and instantaneous effective channel information, it can remove the residual interference after the beamforming without any loss in the rate. The simple BST scheme can only achieve a comparable sum spectral efficiency in the $\text{SNR}_{\text{BBF}}$ range below $-10$ dB. Figure 3.7(b) shows the result if the AoA/AoD information is not perfect but estimated by a BA co-simulation. The BA co-simulation runs with the simulation parameters (i.e., the channel parameters and the pseudo-noise sequence length) as specified in Section 3.2.1 with $\kappa_u = 4$, $\kappa_v = 8$, and a total radiated power constraint. The BA performance can be seen in Figure 3.8. The data communication simulation uses the estimated AoA/AoD after $T_{\text{BA}} = 64$ training slots. For $\text{SNR}_{\text{BBF}}$ values below $-9$ dB, the probability for the correct estimation is not close to one but less. Following, the sum spectral efficiency for this range in Figure 3.7(b) is different than in the ideal case in Figure 3.7(b). Especially, the BZF scheme has a much lower spectral efficiency and performs even worse than the BST scheme. Nonetheless, in a well-designed system, the detection probability of the BA should be close to one and the

[11] Coherent communication is a standard feature of such communication systems and can be achieved using well-known methods, e.g., pilot symbols per user.




(a)




(b)

Figure 3.7.: The achievable ergodic sum spectral efficiency of $K = 4$ users over the $\mathrm{SNR}_{\mathrm{BBF}}$ for the data communication phase with different precoding schemes: (a) with perfect knowledge of the AoAs and AoDs, and (b) in a co-simulation with the BA and, hence, non-perfect AoA/AoD information.

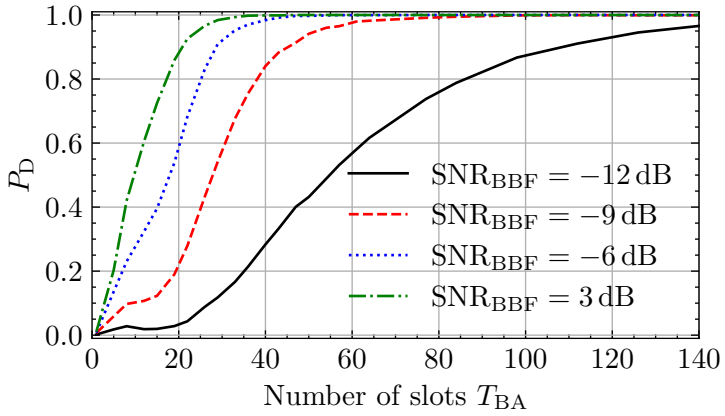



Figure 3.8.: Detection probability $P_D$ of the co-simulation for the data rate analysis of the advanced BA scheme with $M = 64$, $N = 8$, $\kappa_v = 4$, $\kappa_u = 8$, $M_{RF} = 4$, a total radiated power constraint, and multiple $SNR_{BBF}$ values.

BZF scheme would outperform the BST scheme. In such a system, the estimating UE could increase the BA performance by estimating the BA measurement SNR and appropriately changing its codebook, e.g., the $\kappa_v$ value. In conclusion, the BST scheme is only a suitable solution if the digital signal processing has to be very simple (e.g., because the processing capabilities are minimal) or the instantaneous effective channel can not be measured.

### 3.2.3. Effect of the Structure of the Analog Network

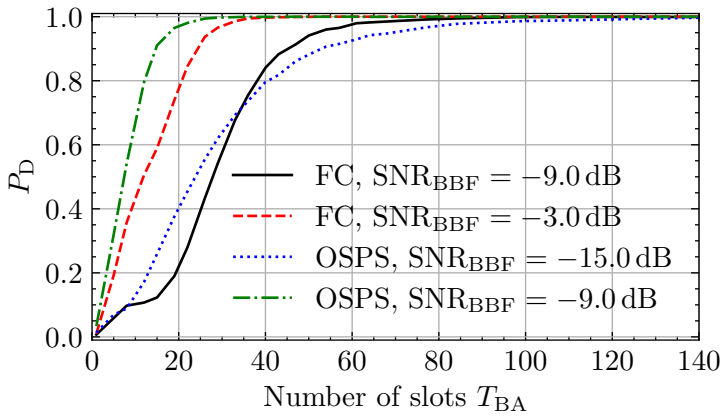
The structure of the analog network influences the signal processing performance of the BA phase and the data communication phase. The antenna array layout and the structure of the network are closely related. We continue assuming a ULA. The layout of a ULA depends only on the number of elements. The number of elements is fixed when comparing the two structures. Hence, the whole antenna configuration is fixed and does not influence the comparison of the structures. Other array layouts would introduce additional parameters, which would make the comparison less general. For the OSPS structure, we assume, that the sub-arrays are so close to each other, that they all see the users at the same azimuth angle. In the ideal case, which we simulate, the sub-arrays overlap each other. In Chapter 5, we simulate a non-ideal antenna array where the sub-arrays are stacked on top of each other. We do not see any deviation in the performance.

The plots in this section have been generated with a numerical simulation. We simulated two systems with $M = 64$ antenna elements and $M_{RF} = 4$ RF chains. One with an FC structure and one with an OSPS structure. The channel parameters and all other parameters are as defined for the simulations in Sections 3.2.1 and 3.2.2. Both systems

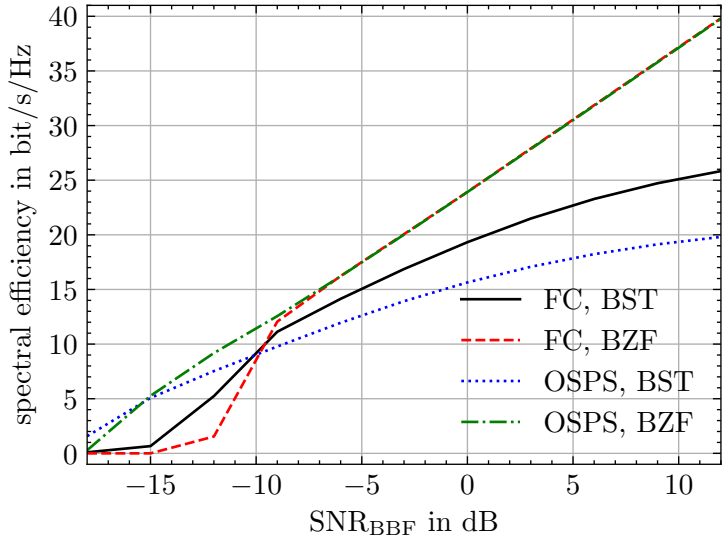
have a total radiated power constraint and have the same $P_{\text{tot}}$. The system based on an FC analog network uses patterns with $\kappa_u = 8$. The OSPS structure-based system employs $\kappa_u = 4$. These values are optimized to give the best BA performance.

One important difference between the two structures is the number of antenna elements that are used to steer beams towards a UE. This number is equivalent to the number of elements connected to each RF chain. In the FC structure, $M$ elements are connected to each RF chain. Whereas in the OSPS structure, $M/M_{\text{RF}}$ elements are connected to each RF chain. The beamforming pattern calculation depends on the number of steerable elements. In the FC case, the Fourier transformation matrix $\mathbf{F}_{\text{BS}}$ as defined in (3.6) depends on $M$. In general, $\mathbf{F}_{\text{BS}}$ depends not on the overall number of elements but the number of steerable elements. Thus, it is different for the two structures. $\mathbf{F}_{\text{BS}}$ defines the grid on which both the BA estimation and the hybrid precoding work. For the BA phase, the AoDs are estimated on a grid of $M$ points for the FC structure and a grid of $M/M_{\text{RF}}$ points for the OSPS structure. Thus, the angular resolution of the AoD estimation is much finer with the FC structure and also the number of possible grid points is much larger. We define the successful estimation for the FC structure on the $M$ point grid and for the OSPS structure on the $M/M_{\text{RF}}$ point grid. One might argue that this is an unfair comparison, but the AoD is estimated to be used in the data communication and only required on the relevant grid. It stands to reason, that due to the much larger search space for a system with an FC analog network the BA requires a longer training time $T_{\text{BA}}$. Figure 3.9(a) shows the detection probability of the successful BA estimation for the two simulated systems for different $\text{SNR}_{\text{BBF}}$ values. The OSPS system requires only $\sim 20$ training slots to achieve $P_{\text{D}} > 0.95$ when the $\text{SNR}_{\text{BBF}}$ is high (in this case above $-9\,\text{dB}$). The FC system reached $P_{\text{D}} > 0.95$ with a high $\text{SNR}_{\text{BBF}}$ (in this case above $-3\,\text{dB}$) after $\sim 30$ slots. So indeed, the OSPS structure requires much less training time compared to the FC structure. It is also interesting to see, that the required $\text{SNR}_{\text{BBF}}$ to achieve a working BA is much lower for the OSPS structure. Nonetheless, both structures successfully estimate the AoAs/AoDs after a small number of training slots. As in Section 3.2.2, we use the estimation result after 64 training slots for the data communication simulation.

The MU-MIMO gain of the data communication mainly depends on the ability to cancel the interference between the users. In a system using single-beam patterns, the level of the interference does depend on the beam width and the level of the side-lobes of the pattern. The beam width decrease with a larger number of antenna elements. If a beam is too wide, it could cover multiple users which have a similar AoD. This problem can be neglected because we assume a working spatial scheduler that schedules users who are sufficiently separated in the angular domain. Otherwise, the OSPS structure with the coarser estimation grid and the broader beams due to the lower number of steerable




(a)




(b)

Figure 3.9.: Comparison of the FC structure and the OSPS structure for a system with $M = 64$, $N = 8$, and $M_{RF} = K = 4$: (a) The detection probability $P_D$ for different $SNR_{BBF}$ values. With $\kappa_v = 4$, $\kappa_u = 8$ for the FC case, and $\kappa_u = 4$ for the OSPS case. (b) The achievable ergodic sum spectral efficiency over the $SNR_{BBF}$ for both precoding schemes.

elements would be at a disadvantage. The side-lobe level of the used single-direction beams, in general, does not depend on the number of elements [54, Ch. 22]. Still, if the number of array elements increases, the highest side-lobes become more narrow and are located closer to the main-lobe in the angular domain. Hence, the energy of the transmission is focused in a smaller angular region for a larger number of steerable elements. This suggests, that the beamforming with the FC structure creates less interference between the users than the OSPS structure. Figure 3.9(b) shows the sum spectral efficiency of the two structures and the two digital precoding schemes. Due to the very good BA performance even down to $\mathrm{SNR}_{\mathrm{BBF}} = -15\,\mathrm{dB}$, the OSPS structure with the BZF precoding achieves high spectral efficiency over the whole SNR range. The FC structure achieves the same spectral efficiency above $\mathrm{SNR}_{\mathrm{BBF}} = -9\,\mathrm{dB}$ where the BA does reliably estimate the AoDs. For both structures, the BST precoding is worse than the BZF precoding. The gap increases as the SNR increases. The performance of the BST precoding and the OSPS structure is worse than that of the BST precoding and the FC structure. These results support the hypothesis, that the larger the number of steerable elements is, the lower is the interference after the beamforming. Anyhow, the zero-forcing precoder in the BZF scheme can successfully cancel the residual interference even in the OSPS case.

In summary and even without considering the beneficial lower complexity of the OSPS structure, it outperforms the FC structure in the BA performance and the data communication performance when the BZF digital precoding is used.

## 3.3. Analog Hardware

This section will investigate the analog hardware aspects of the HDA architecture. The key HDA-specific part of a hybrid system is the analog network. Other analog parts, like the RF chain or the antenna array, are out of the scope of this thesis. The analog hardware has an impact on the signal processing performance of the overall hybrid system but also on other characteristics, e.g., the complexity of the power efficiency. The structure of the analog network is a hardware-related property but was already discussed in Section 3.1 because of the close relation to the signal processing. The impact of the structure on the power efficiency and the hardware complexity is part of this section.

### 3.3.1. Components of the Analog Network

As already introduced in Section 3.1, the main hardware components of the analog network are the power combiners/dividers and the vector modulators. In this section, the characteristics and parameters of these components are covered. We consider phase shifters and

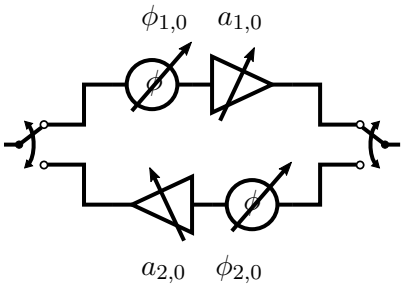



Figure 3.10.: The block diagram of a non-reciprocal analog network.

amplitude modulators combined as vector modulators. Some HDA systems use only phase shifters and no amplitude modulators. This restricts the type of beamforming patterns which the analog network can create. For example, such a system could not create the multi-directional patterns required by the described BA algorithm.

First, we want to consider the reciprocity property of the analog network. Many hybrid systems require the analog network to be reciprocal. A reciprocal network can create the same beamforming matrix $\mathbf{U}_{\mathrm{RF}}$ in the transmit and receive direction up to a constant scalar. Mathematically, $\mathbf{U}_{\mathrm{RF}}^{\mathrm{TX}} = \alpha_{\mathrm{r}}\mathbf{U}_{\mathrm{RF}}^{\mathrm{RX}}$, where $\alpha_{\mathrm{r}} \in \mathbb{C}$ is a constant scalar. Reciprocity is required if a hybrid system uses the physical channel reciprocity, e.g., during the BA or to estimate the effective channel $\tilde{\mathbf{H}}$. Even the AoA/AoD information from the BA phase could be different between the uplink and the downlink if the analog network is not reciprocal. This follows from the fact, that the estimation is based on the known codebook $\mathcal{C}_{\mathrm{T}}$ which includes the exact knowledge of $\mathbf{U}_{\mathrm{RF}}$. Reciprocity can be guaranteed in two ways. First, the physical components are reciprocal (hence, the complete network is reciprocal). Signals traveling through the network in both directions undergo the same transformation (e.g. phase shift). Second, the network is calibrated to be reciprocal. The controllable components of a hybrid analog network can be used to calibrate a non-reciprocal network. Figure 3.10 shows the block diagram of a non-reciprocal network that uses TDD switches. $a_{1,0}$, $\phi_{1,0}$, $a_{2,0}$, and $\phi_{2,0}$ are the amplitude/phase values at control word 0 for both paths, respectively. $a_{1,0} \neq a_{2,0}$ and $\phi_{1,0} \neq \phi_{2,0}$ in an uncalibrated network. During a calibration measurement, the difference $\phi_{\mathrm{c}} = \phi_{1,0} - \phi_{2,0}$ and $a_{\mathrm{c}} = a_{1,0}/a_{2,0}$ between the two paths would be measured. The calibrated network would use these measurements and set the phase and amplitude, e.g., of path 1 to be $\phi'_{1,0} = \phi_{1,0} - \phi_{\mathrm{c}}$ and $a'_{1,0} = a_{1,0}/a_{\mathrm{c}}$. The resulting network would be reciprocal and could be used in a TDD hybrid system.

### Power Dividers and Power Combiners

Power dividers/combiners are the components that span the tree of the network and distribute/accumulate the analog signals to/from the vector modulators. In the following, we consider power combiners/dividers which have three ports (e.g., combining two into one). Combiners/dividers with a higher number of ports and a higher port ratio are built cascading multiple smaller combiners/dividers in a tree structure. For example, a nine-port power combiner that combines eight ports into one can be built with seven three-port power combiners.[12] Power combiners/dividers can be either passive (i.e., they have no power gain) or active (i.e., they have a power gain). Nearly all hardware designs of power dividers/combiners for analog beamforming networks are passive.

In general, a passive three-port component (in the hardware literature also called a three-port network) can only have two of the following three properties: it is reciprocal, it is lossless, and it is matched [55, Ch. 7].[13] Components in RF circuits like the analog network are usually required to be matched. The most common matched passive power divider/combiner is the so-called Wilkinson divider. It is reciprocal and not lossless [55, Ch. 7]. A Wilkinson power divider is due to the reciprocity also a power combiner. A Wilkinson divider can be implemented as a transmission line structure on a printed circuit board or in an integrated circuit. Alternatively, it is commercially available as a circuit component which is usually optimized with respect to its size and bandwidth using different materials. Since a Wilkinson divider is not lossless, it has an $\alpha_{\text{com}}/\alpha_{\text{div}}$ as defined in (3.1) which is equal to its port ratio, e.g., $\alpha_{\text{com}} = 1/2$ for a two to one combiner. Different matched, reciprocal but not lossless power divider concepts can be found in [55, Ch. 7].

A second option to implement power combiners/dividers is the use of amplifiers in integrated circuits. Amplifiers are active non-reciprocal components. Such active combiners/dividers can not only span the tree of the analog network but do also amplify the signals. Thus, $\alpha_{\text{com}}/\alpha_{\text{div}}$ of active combiners/dividers can also be larger than 1. The authors of [62] have designed an analog beamforming receiver for an HDA system using active combiners. Again, such non-reciprocal networks require switches and calibration to be suitable for typical hybrid systems.

As we can see, different power combiner/divider types have different values for $\alpha_{\text{com}}/\alpha_{\text{div}}$. $\alpha_{\text{com}} < 1$ or $\alpha_{\text{div}} < 1$ represent a power loss in the analog network. This power loss needs to be compensated by amplification within, in front, or after the network. From a hardware design perspective, this means an increase in power consumption and circuit complexity. Although, the exact increase of the power consumption depends highly on

[12] Such a tree structure has three levels (eight-to-four, four-to-two, and two-to-one) with four, two, and one three-port power combiners per level.

[13] A lossless network has a power gain of one. A matched network does not reflect signal power on its ports.

the implementation of the amplification and can not be modeled in a general way. From a communication perspective without loss of generality, the factors $\alpha_{\text{com}}$ and $\alpha_{\text{div}}$ can be neglected [27]. In the signal processing analysis, we can set $\alpha_{\text{com}} = 1$ and $\alpha_{\text{div}} = 1$ due to the power constraint and the corresponding normalization of $\mathbf{U}_{\text{RF}}$.

### Vector Modulators

The controllable beamforming of the analog network is realized by one vector modulator per path of the network. As already stated in Section 3.1, we are considering analog networks with phase and amplitude control. Such networks use vector modulators. Most vector modulators, on the other hand, are a combination of a phase shifter and an amplitude modulator (e.g., see the commercial beamforming ICs listed in Section 3.1). In the following, we discuss the properties of phase shifters and amplitude modulators.

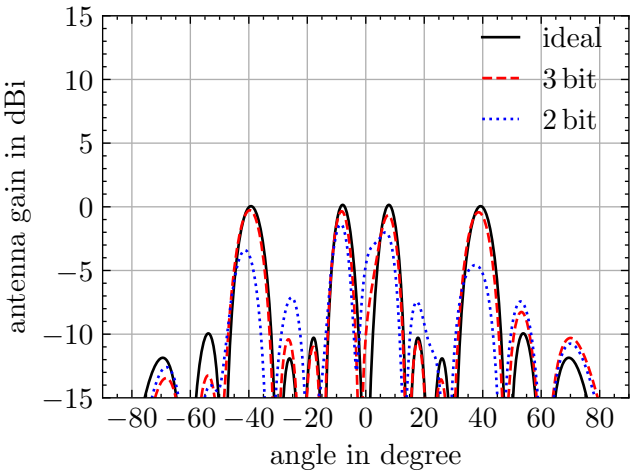
A phase shifter changes the phase of a signal passing through it. It can change the phase within a phase range. The phase range has to cover ±180° or 360° to not restrict the beamforming capabilities of the analog network. Nearly all commercial and scientific designs are capable of covering 360°. Phase shifters can either be implemented to change the phase continuously over the range or on a discrete grid. Since phase shifters are controlled from a digital (discrete) system, even the continuous type is configured on a discrete grid. The resolution of the grid is one of the main parameters of a phase shifter. The resolution can be either defined by the step size $\Delta\phi$, the total number of steps $n_\phi$, or the bit width of the control word $b_\phi$. For a phase range of 360°, the relation between the step size and the bit width is $\Delta\phi = \frac{360°}{n_\phi - 1} = \frac{360°}{2^{b_\phi} - 1}$. Due to the non-ideal behavior of the hardware, a phase shifter has a phase error parameter. This parameter specifies the error between the configured and the real phase change. It is usually in the range of $\pm\Delta\phi/2$ and can vary over the control range. A phase shifter has also a certain phase offset and amplitude offset (i.e., an insertion loss). The phase and amplitude offsets are the static phase and amplitude deviations when the phase is set to be 0°.

Phase shifters can either be implemented to change the phase of the signal or the delay of the signal. The classical phased array processing does assume the signal to be narrowband and, hence, requires a phase difference between the elements of the array [63]. Without this narrowband assumption, the difference between the signals of the antenna elements is a constant delay and not a constant phase. Following, for wideband systems the analog network needs to use time delay shifters, also called true time delay phase shifters. "Standard" phase shifters result in wider beams if the narrowband assumption is not met [63]. This effect is called beam squinting. Thus, the bandwidth of a phase shifter can be either limited by its hardware design or by the narrowband assumption and how

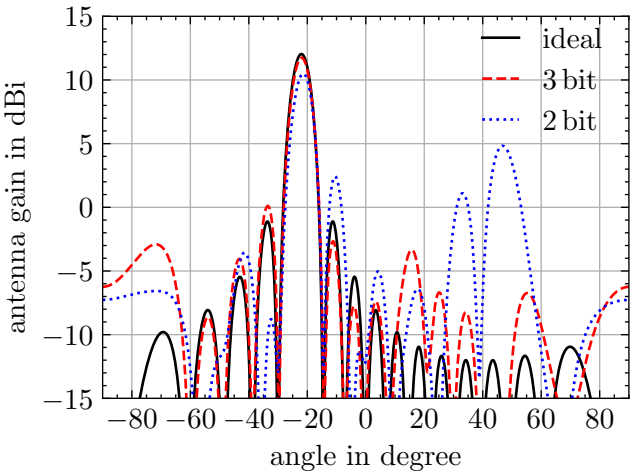
much beam squinting is tolerable. Phase shifters often support a larger range for the center frequency of the signal with a smaller signal bandwidth around the center.

The amplitude modulator changes the amplitude of the signal. The amplitude-related parameters of an amplitude modulator can be given in the linear or the logarithmic (as dB) scale. Both scales can be converted into each other. We will use the logarithmic scale. The range of the amplitude modulator defines the maximum and the minimum of the amplitude change. A typical range is $-30\,\mathrm{dB}$ to $0\,\mathrm{dB}$. An amplitude modulator can either include an amplification (with the maximum of the range being $>0\,\mathrm{dB}$) or only attenuation (with the maximum of the range being $\leq 0\,\mathrm{dB}$). Amplitude modulators including amplification are often implemented using variable gain amplifiers (VGAs). Amplitude modulators have, similar to phase shifters, a control resolution with a specified step size $\Delta a$, number of steps $n_a$, or bit width of the control word $b_a$. The quantization of the range can either be in the linear or the logarithmic scale from which the logarithmic scale is more common. For example, a 4-bit amplitude modulator with a range of $-30\,\mathrm{dB}$ to $0\,\mathrm{dB}$ which is quantized in the logarithmic scale has a step size of $2\,\mathrm{dB}$. The difference between the selected amplitude factor and the realized amplitude factor is the amplitude error which can vary over the control range and should be within $\pm\Delta/2$. Each amplitude modulator has a certain phase offset. The phase offset is the phase change to the signal passing through the amplitude modulator.

Most hybrid signal processing algorithms, including the ones used in this thesis, require explicit beam patterns. This requires the elements of $\tilde{\mathbf{U}}_{\mathrm{RF}}$ to match the configured and in hardware realized phases and amplitudes of the analog network. The non-ideal hardware of the analog network does result in phase and amplitude errors for each path of the network. These errors break the relation between the selected $\tilde{\mathbf{U}}_{\mathrm{RF}}$ and the actual beamforming. The errors can be constant or varying with respect to the phase and amplitude control. For example, differences in the lengths of the paths result in a constant phase error. The phase shifters and amplitude modulators can have varying errors over their control range. The amplitude offset of a phase shifter can vary over the phase control and the phase offset of the amplitude modulator can vary over the amplitude control. Also, the phase and amplitude control of the phase shifter and amplitude modulator, respectively, can deviate from the given quantization grid. All errors need to be corrected up to a certain tolerance to enable the analog network to be used for the described algorithms. This correction can be based on a calibration measurement. The calibration can be used to correct the static errors, the varying errors, and also the reciprocity errors as described above. An investigation of calibration schemes is not part of this thesis. To the knowledge of the author, even the literature does not cover many calibration schemes for hybrid systems.




(a)




(b)

Figure 3.11.: Effect of the resolution of the vector modulator on the beamforming pattern of a 16-element array. The lines show the antenna array gain with no vector quantization (ideal), 3-bit resolution, and 2-bit resolution. (a) Gain of a 4-direction BA beam pattern. (b) Gain of a single beam pattern.

The resolution of the phase shifters and the amplitude modulators affect the beamforming pattern. Figure 3.11 shows exemplary beamforming patterns of a 16-element antenna array using multiple phase and amplitude resolutions. We compare a 2-bit resolution, a 3-bit resolution, and an ideal case without any quantization. Figure 3.11(a) is the array gain of a 4-direction pattern as used during the BA phase. Such patterns depend on the phase and amplitude. Figure 3.11(b) shows the array gain of a single beam pattern as used during the data communication phase. Single beam patterns only require phase modulation. One can see, that the non-ideal resolution does not change the direction of the beam patterns but the gain of the main-lobe and the side-lobe levels. For such a 16-element array, a resolution of 3 bits does barely change the gain in the main-lobe and keeps the side-lobes 12 dB below the main-lobe. In contrast, a 2-bit resolution does severely change the pattern by changing the position of the side-lobes and increasing their gain. This is true for both beam pattern types.

*Numerical simulation of the signal processing:* to determine the effect of the non-ideal beam patterns (so the bit width of the phase and amplitude control) on the signal processing performance, we run a numerical simulation of the BA algorithm and the hybrid precoding algorithm. The simulation is based on the same simulation environment as described in Section 3.2.2. We use the same channel model and system parameters (three multipath components, same strengths, same angular user separation, and so forth). We simulate two systems, both systems have $M = 64$ antenna elements and $M_{\mathrm{RF}} = K = 4$ users and RF chains. One system has an FC analog network and the second system has an OSPS analog network. We simulate the systems with different phase and amplitude resolutions (2 bits to 6 bits) of the beamforming coefficients. In addition, a random error value is added to each phase and amplitude value to represent hardware tolerances and calibration errors. The error value is drawn from a uniform distribution between plus and minus half of the resolution. The overall phase range is $\pm 180°$. For example, a 4-bit resolution results in a quantization of the phase with 24° and a uniform random error of $\pm 12°$. The amplitude range is $-30$ dB to 0 dB and the amplitude is quantized in the logarithmic scale. In the 4-bit example, the amplitude is quantized with 2 dB and the uniform random error is bounded by $\pm 1$ dB. As a benchmark, the system is also simulated with ideal phase and amplitude settings (no quantization).

Figure 3.12 shows the effect of the resolution on the detection probability $P_{\mathrm{D}}$ of the BA. Figures 3.12(a) and 3.12(b) are the results of the FC system and Figures 3.12(c) and 3.12(d) are the results of the OSPS system. The $\mathrm{SNR}_{\mathrm{BBF}}$ in Figures 3.12(a) to 3.12(d) is $-9$ dB, 3 dB, $-15$ dB, and 3 dB, respectively. In all plots, it is noticeable that a resolution of 4 bits or more does achieve the same performance as the ideal case. Even a resolution of 3 bits achieves very good performance for the higher $\mathrm{SNR}_{\mathrm{BBF}}$ case (Figures 3.12(b) and 3.12(d)) and only

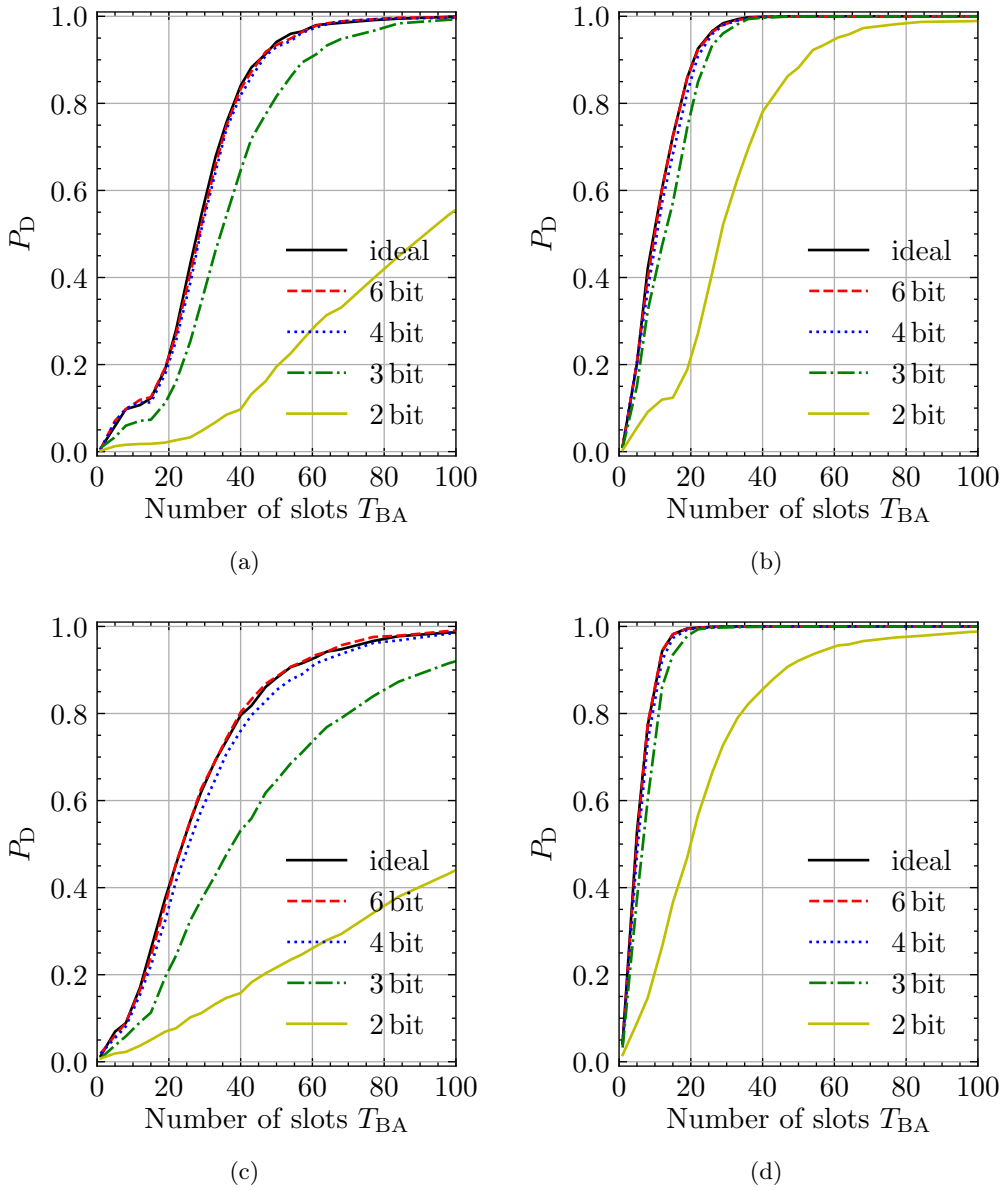



Figure 3.12.: Effect of the resolution of the vector modulator on the detection probability $P_D$ of the BA. The probability is given for the ideal case (no vector quantization) and resolutions of 2 bits to 6 bits. The system has $M = 64$, $N = 8$, $M_{RF} = 4$, $\kappa_v = 4$ in all plots. (a) and (b) are for an FC system and $\kappa_u = 8$. The $SNR_{BBF}$ is $-9$ dB in (a) and 3 dB in (b). (c) and (d) are for an OSPS system and $\kappa_u = 4$. The $SNR_{BBF}$ is $-15$ dB in (c) and 3 dB in (d).

a moderate performance loss for the lower $\mathrm{SNR}_{\mathrm{BBF}}$ case (Figures 3.12(b) and 3.12(d)). Only the 2-bit resolution shows a severe degradation of the BA performance. When we compare the two structures of the analog network, the effect of the bit width on the BA performance is similar for both. This is surprising because one could imagine that the higher angular resolution of the FC network makes it more susceptible to a negative effect of the bit width on the performance. The simulation does not support this presumption.

The results of the spectral efficiency simulation can be seen in Figure 3.13. Figures 3.13(a) and 3.13(b) are the results of the FC system and Figures 3.13(c) and 3.13(d) are the results of the OSPS system. The curves in Figures 3.13(a) and 3.13(c) show the results of the BZF precoding scheme. Similar to the BA results, the performance does not degrade for a phase and amplitude resolution which is equal to or higher than 3 bits. The performance loss of the spectral efficiency due to the 3-bit, 4-bit, and 6-bit resolution can be neglected. In contrast, the performance using a 2-bit resolution does drop significantly over the whole $\mathrm{SNR}_{\mathrm{BBF}}$ range. The performance is degraded even in an $\mathrm{SNR}_{\mathrm{BBF}}$ range $\geq$3 dB where the BA with high probability estimates the correct AoAs/AoDs. This suggests, that the resolution of the vector modulators affects directly the precoding. In contrast to the BZF scheme, the effect of the bit width is much stronger on the simple BST scheme. As one can see in Figures 3.13(b) and 3.13(d), the spectral efficiency of the systems with 2-bit, 3-bit, and 4-bit resolution degrades rapidly the lower the resolution is. Only the system with a 6-bit resolution of the phase and amplitude shows similar performance as the ideal system. The ZF precoder can cancel the interference which is created due to the lower resolution and the non-perfect beam patterns. Comparing the FC structure and the OSPS structure, both show similar behavior with respect to the bit width.

In Figure 3.14, we have analyzed the effect of the bit width on an FC system with a higher number of RF chains. The simulated system has eight RF chains and schedules eight users at the same time. Following, the users are located much denser in the angular dimension than before. We only show the results for the precoding simulation as the BA results are not dependent on the number of users.[14] In Figure 3.14, the system shows a similar effect of the bit width on the performance compared to the systems in Figure 3.13. The BZF scheme for a higher number of users is still robust against the resolution of the vector modulator.

As the beamforming coefficient calculation depends on $M$ (see (3.6)), it stands to reason that a system with a larger number of antenna elements requires a higher phase resolution. Figures 3.15(a) and 3.15(b) show the BA and the data rate performance of a system with

[14] Please note, the performance of the BA still depends on the number of RF chains as described in Section 3.2.1. Nonetheless, the effect of the resolution of the phase and amplitude control is independent of the number of RF chains.

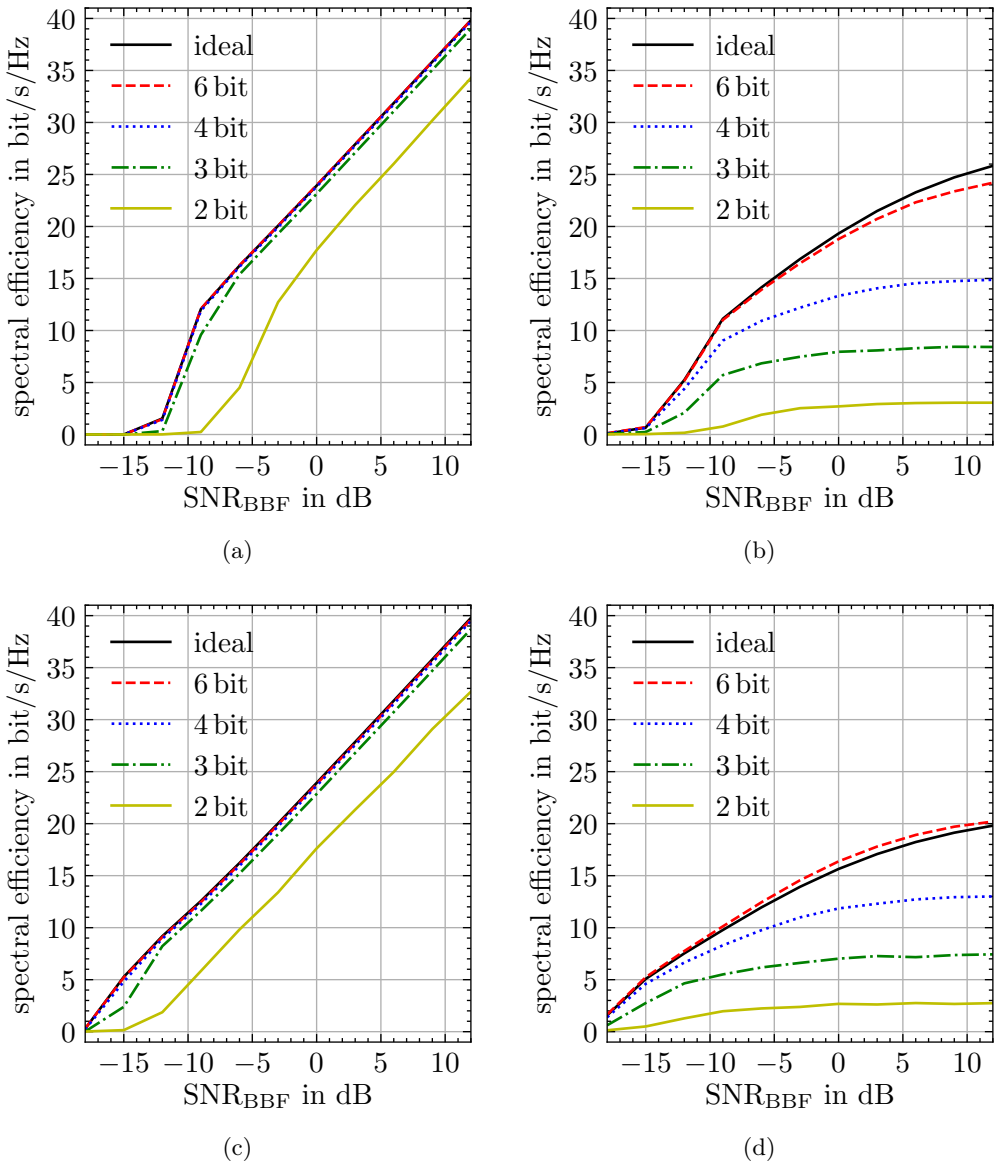



Figure 3.13.: Effect of the resolution of the vector modulator on the asymptotic ergodic spectral efficiency. The spectral efficiency is given for the ideal case (no vector quantization) and resolutions of 2 bits to 6 bits. The system has $M = 64$, $N = 8$, $M_{\mathrm{RF}} = 4$, and was simulated in a co-simulation with the BA in all plots. (a) and (b) are for an FC system and the BZF scheme and the BST scheme, respectively. (c) and (d) are for an OSPS system and the BZF scheme and the BST scheme, respectively.

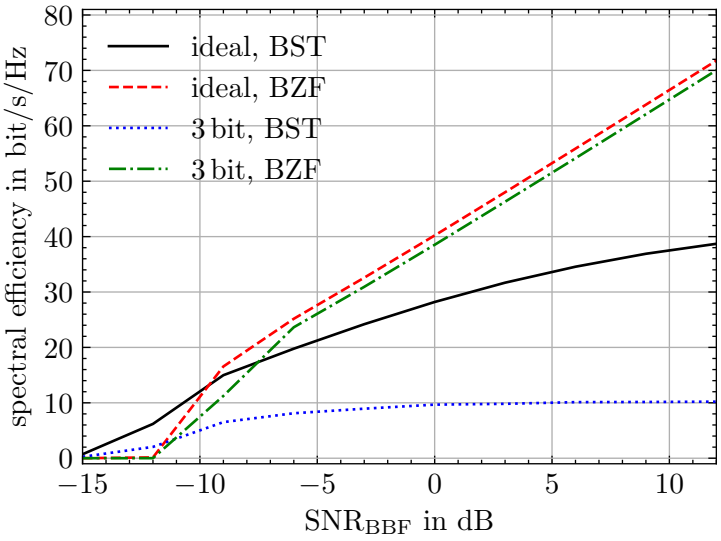



Figure 3.14.: Effect of the resolution of the vector modulator on the asymptotic ergodic spectral efficiency of a system with $M_{\mathrm{RF}} = 8$ RF chains. The system has the same parameters as in Figure 3.12(a) except for $M_{\mathrm{RF}}$. The spectral efficiency is given for the ideal case (no vector quantization) and a 3-bit resolution.

$M = 256$ antenna elements for the ideal case, a 4-bit resolution, and a 3-bit resolution. In comparison to Figure 3.12, the larger system does require more training slots $T_{\mathrm{BA}}$. This stands to reason since the channel dimension is much larger. Nonetheless, if the $\mathrm{SNR}_{\mathrm{BBF}}$ is sufficiently large enough, the BA does estimate the correct AoAs/AoAs for $T_{\mathrm{BA}} = 64$ even with a 3-bit resolution. As can be seen in Figures 3.15(c) and 3.15(d), the precoding performance with a 3-bit resolution does degrade over the whole $\mathrm{SNR}_{\mathrm{BBF}}$ range for both the BST scheme and BZF scheme. The performance of the BZF scheme with a 4-bit resolution does not degrade. Thus, the larger the number of array elements is the larger the bit width should be. A bit width of 3 bits is not sufficient for 256 antenna elements.

In conclusion, we can recommend a minimum resolution of 3 bits for hybrid systems with $M = 64$ antenna elements as long as the precoding is based on the BZF scheme. A larger resolution is required, if $M$ is larger, e.g., 4 bits for $M = 256$. We have not investigated the effect of the bit width of the vector modulators at the UEs. Although, we do not expect different results. Due to the much smaller number of antenna elements, even a 2-bit resolution might likely be sufficient.

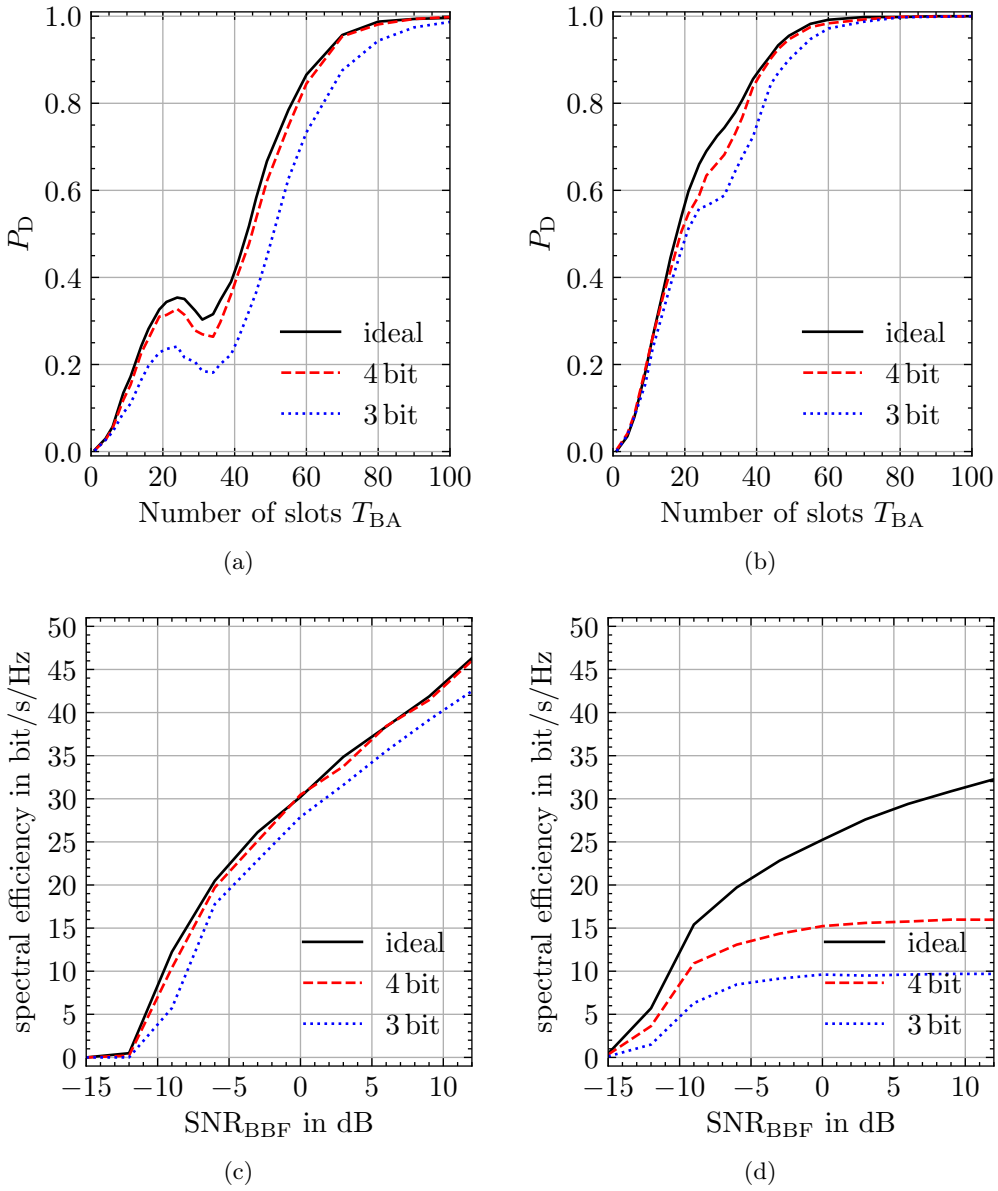



Figure 3.15.: Effect of the resolution of the vector modulator on the performance of a system with $M = 256$ antenna elements. The system has the same parameters as in Figure 3.12(a) except for $M$ and $\kappa_u$. The results are given for the ideal case (no vector quantization) and 3-bit and 4-bit resolution. (a) and (b) show the BA performance with $\kappa_u = 16$, and $\mathrm{SNR}_{\mathrm{BBF}} = -9\,\mathrm{dB}$ and $\mathrm{SNR}_{\mathrm{BBF}} = 3\,\mathrm{dB}$, respectively. (c) and (d) show the ergodic spectral efficiency of the BZF scheme and the BST scheme, respectively.

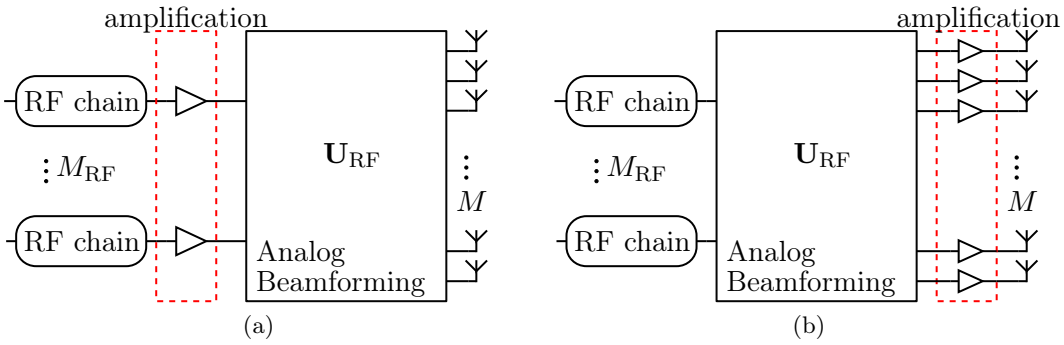



Figure 3.16.: Block diagrams of the possible locations for amplifiers within the TX path of a hybrid system. (a) The amplifiers are part of the RF chains (or in between the RF chains and the analog network). (b) The amplifiers are between the analog network and the antenna elements.

### 3.3.2. Amplifier Considerations

We do not consider amplifiers to be a direct part of the analog beamforming network. Nonetheless, the analog network has an impact on the amplification design of a hybrid system. The main impact is the insertion loss of the analog network which needs to be compensated by amplification. The concrete insertion loss does depend on the actual design of the analog network and can not be given exactly by a simple mathematical formula. The overall insertion loss depends among other things on the power combiner/divider type, the amplitude modulator loss/gain, the insertion loss of the components, and the structure of the network. The overall insertion loss of the analog network should be included in the link budget calculation during the hardware design.

In the following, we discuss some aspects of the analog network/amplification design to understand their impact on the system. As already described in Section 3.1, the structure has an impact on the insertion loss. Especially when considering Wilkinson power dividers for the power distribution/combining network, we can mathematically give the required gain for both structures. We define the required power gain as $A_{\mathrm{req}} = 1/(\alpha_{com} \cdot \alpha_{div})$. Using Wilkinson dividers, the FC structure would require a gain of $A_{\mathrm{req}}^{\mathrm{FC}} = M \cdot M_{\mathrm{RF}}$ and the OSPS structure $A_{\mathrm{req}}^{\mathrm{OSPS}} = \frac{M}{M_{\mathrm{RF}}}$. It is apparent, that the FC structure requires much more gain than the OSPS structure.

Besides the gain, also the required number of amplifiers and the required maximum available output power depend on the analog network. In the following, we only consider the last amplifier in the transmission (TX) case and the first amplifier in the reception (RX) case. The last amplifier before the antenna in a transmitter is usually called a power amplifier (PA) due to the design goal of maximizing the output power. The first amplifier

in a receiver is usually a low-noise amplifier (LNA) due to the optimized noise figure (NF) characteristic. We consider two cases for the location of the amplification within the hybrid system. The amplifiers can either be on the RF chain side of the analog network (Figure 3.16(a)) or between the analog network and the antenna array (Figure 3.16(b)). In the former case, the amplifiers can also be regarded as parts of the RF chains and a system has in total $M_{\mathrm{RF}}$ amplifiers for each the RX and the TX part. In the latter case, a system has in total $M$ amplifiers per RX and TX part. The main amplifier characteristic of the TX PAs is the required output power. The required output power per amplifier depends on the power constraint of the system. We assume a total radiated power constraint. If the amplifiers are located as in Figure 3.16(b), the per amplifier output power to reach the total output power $P_{\mathrm{tot}}$ is $P_{\mathrm{PA,b}} = P_{\mathrm{tot}}/M$. In the case of Figure 3.16(a), the required PA output power increases by the required amplification of the network and $P_{\mathrm{PA,a}} = A_{\mathrm{req}} \cdot P_{\mathrm{tot}}/M$. In conclusion, the case of Figure 3.16(a) has fewer PAs but with a much higher output power requirement. The other case has more PAs but with a lower output power requirement. The required gain per amplifier is the same for both cases.

For the RX part, the most important amplifier characteristic is the NF. Due to the Friis formula for noise [54, Ch. 16], the NF is mainly determined by the first stage in a cascade of components. If the LNA is in between the antenna array and the analog network, the NF is determined by the LNA. If, on the other hand, the LNAs are placed after the analog network, the insertion loss of the network determines the NF of the system. Of course, the latter case is very undesirable since the insertion loss is often much higher than the typical NF of an LNA. Most hybrid systems will use LNAs after the antennas to decouple the power distribution loss from the NF.

Not only the analog network has an impact on the amplification design but also the other way around. The reciprocity of the analog network does, among other things, depend on the location of the amplifiers. Amplifiers are in general not reciprocal. Most RF systems use switches for TDD or frequency filters for FDD to separate the transmit and receive chains/amplifiers. The same switch-based concept as shown in Figure 3.10 can be used for amplifiers. Of course, such a circuit and the amplifiers would require calibration to ensure reciprocity similar to the calibration described in Section 3.3.1. If the amplifiers are part of the RF chains as in Figure 3.16(a), only $M_{\mathrm{RF}}$ amplifiers need to be calibrated for reciprocity. A system with the amplifiers on the antenna array side of the analog network as shown in Figure 3.16(b) requires the calibration of $M$ amplifiers. Similar to the channel estimation, the reciprocity calibration for amplifiers (and other components) on the RF chain side of the network is much more straightforward than for components on the antenna array side. As explained in Section 3.2.1, the estimation of the overall channel $\mathbf{H}$ requires more measurements than the estimation of the effective channel $\tilde{\mathbf{H}}$. This is also true for

the reciprocity calibration. However, the "coherence time" of the calibration, meaning how long the calibration data is valid, is much longer than the channel coherence time. Even a one-time calibration after manufacturing might be sufficient. Although, the exact long-time behavior of amplifiers is not well covered in the literature and the former statement is more a rule of thumb of RF electronic engineers.

### 3.3.3. Power Efficiency

The power efficiency of the overall hybrid system depends among other things on the efficiency of each part but also on the structure of the analog network. This section investigates the impact of the structure on power efficiency. It is based on [36] and [27]. The paper [27] was written by the main author but the research related to the hardware constraints including the power efficiency was done by the author of this thesis. For the following comparison, we do only consider the power consumption of the PAs since they dominate the overall system power consumption. We assume that even for the FC structure the consumed power of the analog network is much smaller than the power consumed by all PAs. The power consumption of the RF chains and the digital signal processing might not be negligible but it does not depend on the design of the analog network and can therefore also be ignored.

We also assume that the power consumption of the PAs and the full network does not depend on the gain but rather on the maximum output power. We ignore the required gain $A_{\text{req}}$ and, hence, also the impact of the power distribution network. We choose the amplification design where a PA is placed at every antenna element. However, the calculation is also correct for the case where the amplifiers are part of the RF chains. All PAs have the same gain, maximum output power $P_{\max}$, and efficiency at maximum power $\eta_{\max}$. For any given element in the array, let $P_{\text{rad}}$ denote the radiated power of the element, and $P_{\text{cons}}$ denote the consumed power by the corresponding PA including both the radiated power and the dissipated power. We model the consumed power with respect to the radiated power following the approach in [26] as

$$P_{\text{cons}} = \frac{\sqrt{P_{\max}}}{\eta_{\max}} \sqrt{P_{\text{rad}}}. \tag{3.18}$$

This approach models typical PAs very well but could be replaced by measurements or simulations of a chosen PA. The effective efficiency for a given radiated power is

$$\eta_{\text{eff}} = \frac{P_{\text{rad}}}{P_{\text{cons}}}. \tag{3.19}$$

In the FC structure, multiple input signals $x_k(t)$ from different RF chains are combined before the PAs. This superposition of signals and the peak-to-average power ratio (PAPR) of the time-domain transmit waveform $x(t)$ let the input power of the $M$ PAs vary over time. The PAPR of a waveform depends on the modulation of the waveform. In particular, an OFDM modulation has a high PAPR value [64]. To avoid non-linear distortion, the system needs to assure that every PA works below its maximum output power at all times. We generally have two options to compare the two analog network structures and the two different modulation schemes:

*Option I:* Both structures utilize the same PA but apply a different input backoff $\alpha_{\text{off}} \in (0, 1]$, such that the peak power of the radiated signal is smaller than $P_{\max}$. As a reference, we denote as $(P_{\text{rad},0}, \eta_{\max,0})$ the parameters of this reference PA under the reference precoding scheme with a power backoff factor $\alpha_{\text{off},0}$ (as illustrated later in this section). For different signal processing scenarios (with certain $\alpha_{\text{off}}$) the effective radiated power and the consumed power read $P_{\text{rad}} = \frac{\alpha_{\text{off}}}{\alpha_{\text{off},0}} P_{\text{rad},0}$ and $P_{\text{cons}} = \frac{\sqrt{P_{\max,0}}}{\eta_{\max,0}} \sqrt{P_{\text{rad}}}$. The transmitter efficiency is given by

$$\eta_{\text{eff}} = \frac{P_{\text{rad}}}{P_{\text{cons}}} = \frac{\sqrt{P_{\text{rad}}} \cdot \eta_{\max,0}}{\sqrt{P_{\max,0}}}. \tag{3.20}$$

*Option II:* We choose to deploy different PAs for different structures, with a maximum output power given by $P_{\max} = \frac{\alpha_{\text{off},0}}{\alpha_{\text{off}}} P_{\max,0}$, where $\alpha_{\text{off}}$ has the same value as in *Option I*. This means that we scale the maximum power of the PA according to the backoff factor of *Option I*. Consequently, the effective radiated power and the consumed power of the underlying PA read $P_{\text{rad}} = P_{\text{rad},0}$ and $P_{\text{cons}} = \frac{\sqrt{P_{\max,0} \cdot \alpha_{\text{off},0}/\alpha_{\text{off}}}}{\eta_{\max}} \sqrt{P_{\text{rad}}}$. The transmitter efficiency is given by

$$\eta_{\text{eff}} = \frac{P_{\text{rad}}}{P_{\text{cons}}} = \frac{\sqrt{P_{\text{rad}}} \cdot \eta_{\max}}{\sqrt{P_{\max,0} \cdot \alpha_{\text{off},0}}} \cdot \sqrt{\alpha_{\text{off}}}. \tag{3.21}$$

*Option I* corresponds to a comparison with a fixed chosen PA whereas *option II* allows comparing the two architectures with matched PA designs.

We have run numerical simulations to compare both structures. We simulated two modulation schemes, single-carrier modulation (SCM) and OFDM modulation. The PAPR and therefore the efficiency with a linear operation of the PAs depends on the modulation and the structure. We consider a system with $M = 64$ antenna elements and $M_{\text{RF}} = 4$ RF chains.

We first assume a reference scenario as the baseline, i.e., the OSPS architecture using a single-carrier modulation. The reference PA has $P_{\max,0} = 36\,\text{dBm}$ and $\eta_{\max,0} = 0.24$. The backoff factor with respect to different waveforms and transmitter architectures can be

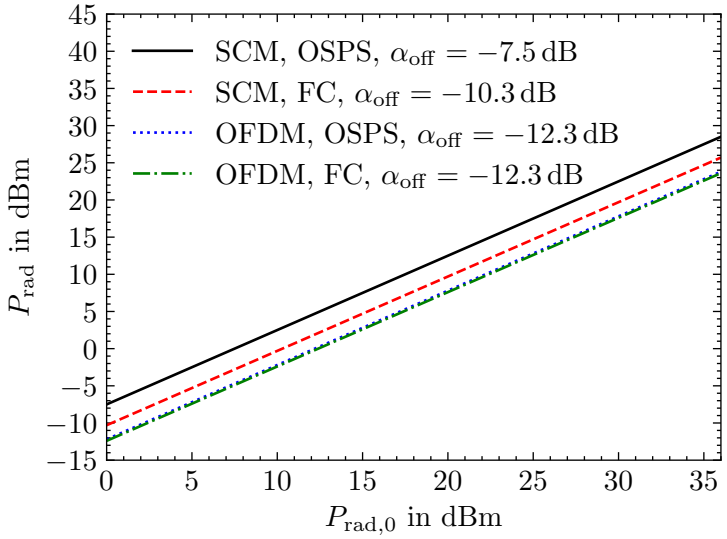
written as $\alpha_{\text{off}} = 1/(P_{\text{PAPR}})$, where $P_{\text{PAPR}}$ represents the PAPR of the input signals at the PAs. The investigation for 3GPP LTE in [64] showed that with a probability of 0.9999, the PAPR of the LTE single-carrier waveform is smaller than ∼7.5 dB and the PAPR of the LTE OFDM waveform (with 512 subcarriers employing quadrature phase-shift keying) is smaller than ∼12.3 dB. We set $P_{\text{PAPR}}$ to these values for the OSPS architecture. For the FC architecture, however, the input signals of the PAs are the sum of the signals from different RF chains. For an OFDM modulation, each signal can be modeled as a Gaussian random process [64] and the signals from different RF chains are independent. The PAPR of the sum of Gaussian random signals is the same as of one of the signals. Therefore, we set the PAPR of the FC structure with OFDM signals also to 12.3 dB. For the case of SCM signaling, there is no clear work in the literature that shows how the sum of single-carrier signals behaves. We simulated the sum of $M_{\text{RF}} = 4$ single-carrier signals using the same parameters as in [64]. The result shows that with a probability of 0.9999, the PAPR of the sum is smaller than ∼10.3 dB. We apply these values and, without loss of generality, we assume $\alpha_{\text{off},0} = 0\,\text{dB}$ as reference.

As can be seen in (3.20), the efficiency with the assumptions of *Option I* only depends on the chosen PA and the radiated power $P_{\text{rad}}$. $P_{\text{rad}}$ itself does not depend directly on the structure but the required backoff $\alpha_{\text{off}}$. Figure 3.17(a) shows the achievable radiated power for the different structures and modulation schemes over the radiated power of the reference scenario (OSPS structure and SCM). The OSPS structure with an SCM results in the highest radiated power followed by the FC structure with an SCM. For the OFDM modulation, both structures have the same efficiency and radiated power which are smaller than the SCM values. Figure 3.17(b) shows the efficiency against the radiated power in the case of *option II*. The difference between the structures and the modulation schemes can be explained by the fact that due to the PAPR of the signals, the PAs cannot achieve their maximal efficiency. Hence, the configuration with an OSPS network and an SCM has the highest efficiency, followed by the FC structure with SCM. Both structures with an OFDM modulation have the lowest efficiency.

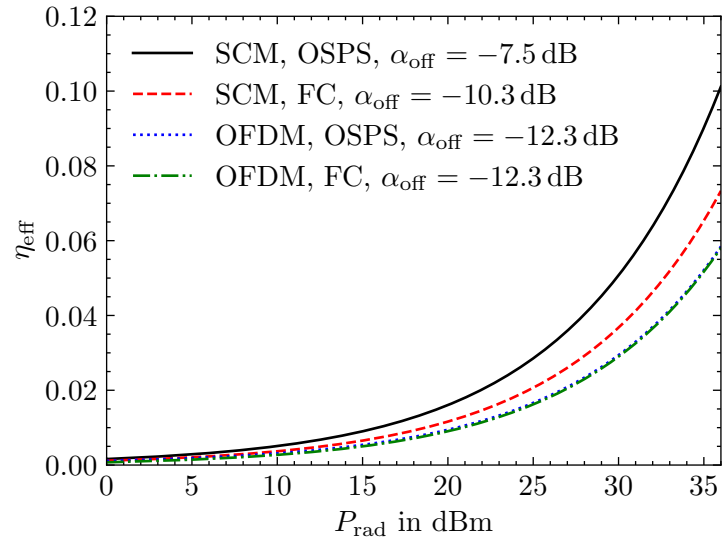
A conclusion of the modeling and the simulations is, that the structure does not directly influence the efficiency of the system (under the given assumptions). Rather, the architecture can change the PAPR of the beamformed signal which in turn impacts the efficiency.

### 3.3.4. Complexity Analysis

As written before, the lower complexity of the HDA architecture compared to the fully-digital architecture for massive MIMO is one of its main advantages [14]. The complexity is a characteristic that is not easy to quantify and which strongly depends on the technology




(a)




(b)

Figure 3.17.: Power efficiency evaluation of the two analog network structures with $M = 64$, $M_{\text{RF}} = 4$, $P_{\text{max},0} = 36\,\text{dBm}$ and $\eta_{\text{max},0} = 0.24$. (a) is the actual radiated power in *option I* over the radiated power of the reference scenario. (b) is the power efficiency in *option II* over the actual radiated power.

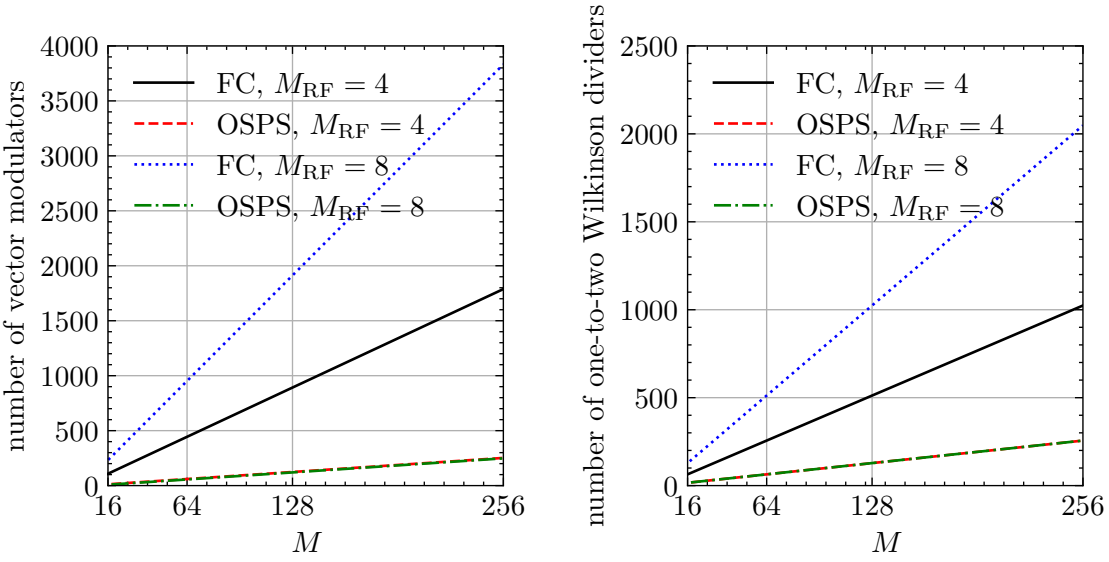



Figure 3.18.: Number of used components for an FC analog network and an OSPS analog network over the number of antenna elements $M$ for $M_{\mathrm{RF}} = 4$ and $M_{\mathrm{RF}} = 8$: (a) number of vector modulators and (b) number of one-to-two Wilkinson power dividers.

or the state-of-the-art. For example, mm-wave devices were not many years ago only used in point-to-point communication links due to their hardware complexity. Whereas in 5G, they are considered to boost the data rate of an average UE. This was made possible by the progress in semiconductor technology. Keeping this in mind, we still want to compare the complexity of the fully-digital architecture with the HDA architecture and the FC structure with the OSPS structure.

We will first compare the FC structure to the OSPS structure. The complexity of a hardware design can be quantified by comparing the number of required components. The main components of the analog network are the power dividers/combiners and the vector modulators. We assume Wilkinson power dividers, although the results might be similar for other power divider types. We consider the number of one-to-two Wilkinson power dividers $N_{\mathrm{WIL}}$ in the power distribution network of the two structures.[15] As described in Section 3.3.1, power dividers with other port ratios can be implemented in a tree structure with one-to-two dividers. $N_{\mathrm{WIL}}$ of the FC structure is given by $N_{\mathrm{WIL}}^{\mathrm{FC}} = (M-1)M_{\mathrm{RF}} + (M_{\mathrm{RF}}-1)M$. The OSPS structure uses $N_{\mathrm{WIL}}^{\mathrm{OSPS}} = M - M_{\mathrm{RF}}$ Wilkinson dividers. Additionally, the number

[15] $N_{\mathrm{WIL}}$ includes all dividers and combiners in the whole analog network. A Wilkinson divider can also be used as a power combiner.

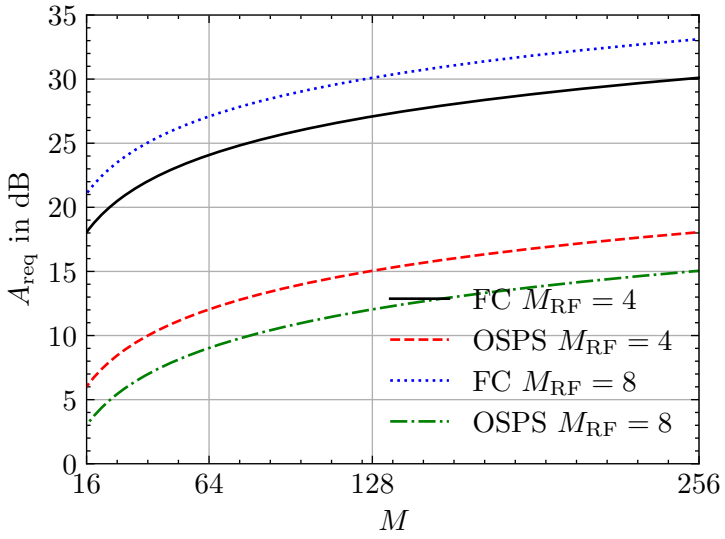



Figure 3.19.: Required gain to compensate for the loss of the power distribution in the analog network for both structures over the number of antenna elements $M$ for $M_{\text{RF}} = 4$ and $M_{\text{RF}} = 8$.

of vector modulators used in the analog network $N_{\text{VM}}$ is different for the two structures. The FC structure has $N_{\text{VM}}^{\text{FC}} = M \cdot M_{\text{RF}}$ vector modulators which is $M_{\text{RF}}$ times more than the OSPS structure with $N_{\text{VM}}^{\text{OSPS}} = M$ vector modulators. For both component types, the FC structure uses much more components than the OSPS structure. The difference increases linearly with the number of antennas $M$. To illustrate the rapid increase, we plot in Figures 3.18(a) and 3.18(b) the number of used vector modulators and Wilkinson power dividers, respectively. The number of used components in the FC structure can become very large even for a moderate number of elements, like 256, and eight RF chains. It is apparent, that the OSPS structure is much better suited for systems with a large number of antenna elements and RF chains.

Besides the analog network by itself, also the required amplification influences the complexity. As written before, the insertion loss of the analog network has to be compensated by amplification. The required gain $A_{\text{req}}$ as given in Section 3.3.2 is different for the two structures. A higher gain requires either bigger or more amplifiers which increases the general complexity of the system. Figure 3.19 shows the required gain over the number of antenna elements for both structures and four and eight RF chains. The gain is plotted in the logarithmic scale in dB. Similar to the number of components, the FC structure also requires a larger gain than the OSPS structure. However, the gap between the two structures with the same $M_{\text{RF}}$ is constant due to the logarithmic scale. The required gain of the FC structure increases with a larger number of RF chains. In contrast, $A_{\text{req}}^{\text{OSPS}}$ decreases

for a larger $M_{\mathrm{RF}}$. In an FC network, the gain has to compensate for the power combiner loss. Whereas in an OSPS network, the power combiner loss is one and the power divider loss decreases when the number of RF chains increases. The power of each RF chain is transmitted over fewer elements. Hence, the OSPS structure is better suited for systems with many RF chains.

In addition to the comparisons above, the hardware development complexity increases with the FC structure. The hardware development complexity can not easily be expressed in a formula. One aspect where the higher complexity can be recognized is the number of signal crossings in the analog network. Signal crossings are not realized without difficulty in RF hardware. In Figure 3.3, it is easy to see that the FC network has signal crossings (crossing lines in the block diagram), whereas the OSPS structure has none. The signal crossings in the FC structure are located in the power combining part of the network. The number of crossings increases with an increasing number of RF chains.

In conclusion, the complexity of the FC structure is much higher compared to the OSPS structure. The number of required components, the required gain, and the development complexity are higher for the FC structure. In addition, the complexity of the OSPS structure scales better when the number of antenna elements and the number of RF chains increases. After comparing the system performance in Section 3.2.3, the power efficiency in Section 3.3.3, and the complexity, the OSPS structure outperforms the FC structure in all domains.

We will now compare the complexity of a hybrid massive MIMO system using an OSPS analog network with a fully-digital system. In Section 2.4, we already explained that the main difference in complexity is the reduced number of RF chains in a hybrid system in comparison to a fully-digital system. A fully-digital system has $M$ RF chains. An HDA system has $M_{\mathrm{RF}}$ RF chains and should be operated with $K = M_{\mathrm{RF}}$ simultaneously scheduled users. $K = M_{\mathrm{RF}}$ maximizes the performance over the cost/complexity ratio. This implies, that the hybrid system should be optimized for the network in which it is deployed. The hybrid approach achieves the lower number of RF chains by employing the analog network. The analog network increases the complexity compared to the fully-digital approach. The complexity of the analog network and the complexity of an RF chain can not easily be compared. We formalize the comparison by considering an abstract complexity value. $C_{\mathrm{FD}}$ is the abstract complexity of a fully-digital system and given by

$$C_{\mathrm{FD}} = C_{\mathrm{RF}} \cdot M, \tag{3.22}$$

where $C_{\mathrm{RF}}$ is the abstract complexity of a single RF chain. The abstract complexity $C_{\mathrm{HDA}}$ of a hybrid system with an OSPS network is

$$C_{\mathrm{HDA}} = C_{\mathrm{RF}} \cdot M_{\mathrm{RF}} + C_{\mathrm{VM}} \cdot M + C_{\mathrm{WIL}} \cdot (M - M_{\mathrm{RF}}) + C_{\mathrm{amp}}, \tag{3.23}$$

where $C_{\mathrm{VM}}$, $C_{\mathrm{WIL}}$, and $C_{\mathrm{amp}}$ are the abstract complexities per vector modulator, Wilkinson power divider, and for the additional required amplification, respectively.

Of course, the different complexities depend among other things on the center frequency, the bandwidth, and, in general, the technology. One has to compare the fully-digital architecture and the HDA architecture for a specific application. A parameter like the component price or the power consumption can be used as the abstract complexity.

As a simple example, we compare the cost of a system operating in the frequency range 24.25 GHz to 27.50 GHz which is one of the mm-wave 5G bands. We assume a bandwidth of 400 MHz which is the largest currently specified bandwidth in the 5G standard [60]. The system has $M = 64$ antenna elements and $M_{\mathrm{RF}} = 8$ RF chains. To simplify the comparison, we represent the cost of an RF chain by the cost of the ADC and DAC. We ignore the cost for other components, e.g., the frequency converters. We use the cheapest ADC and DAC from the manufacturer Analog Devices, Inc.[16] We consider ADCs/DACs with two channels, for IQ sampling, and a minimum resolution of 8 bits. We use the online product selector of Analog Devices, Inc.[17]. The cheapest ADC is the device AD9286 with a price of 43.20 USD. The cheapest DAC is the device AD9780 with a price of 20.02 USD. The total cost per RF chain is 63.22 USD. For the hybrid system, the cost of the analog network is represented by the cost of an integrated beamforming chip. Such circuits include the vector modulators, the power distribution network, and the amplifiers. We have already listed a selection of beamforming ICs in Section 3.1. The only IC which has a public price is the NXP MMW9004KCAZ.[18] It is a four-channel beamforming circuit for RX and TX for the considered frequency band. It includes a one-to-four power distribution network and 24 dB power gain. The NXP MMW9004KCAZ has a price of 50.05 USD.[19] We ignore the cost of the rest of the power distribution network. The cost per antenna element of the hybrid

[16] Analog Devices, Inc. is a large semiconductor company that designs and sells many components. We use it as an example due to its large number of available ADCs and DACs. Thus, we get a realistic comparison of the price of commercially available components. Of course, this is also a momentary comparison.

[17] ADC search: `https://www.analog.com/en/parametricsearch/10826#/` Retrieved 22 December 2021
DAC search: `https://www.analog.com/en/parametricsearch/10956#/` Retrieved 22 December 2021

[18] `https://www.nxp.com/products/radio-frequency/rf-power/rf-cellular-infrastructure/5g-mmwave/24-25-27-5-ghz-4-channel-analog-beamforming-integrated-circuit:MMW9004KC` Retrieved 22 December 2021

[19] `https://www.avnet.com/shop/us/products/nxp/mmw9004kcaz-3074457345645366850?` Retrieved 22 December 2021

system is 12.51 USD. Following (3.22), the total component cost of the fully-digital system is 4046.08 USD. According to (3.23), the hybrid system with an OSPS structure has a total component cost of 1306.56 USD. The cost of the fully-digital system is around three times higher than the cost of the hybrid system. This simple example already shows the big difference in the cost, which is equivalent to the complexity. Please keep in mind, that the example ignores many aspects, e.g., the development cost and many of the components. The assumptions are most likely in favor of the fully-digital system and, hence, do not influence the essential outcome.

## 3.4. Digital Hardware

This section tries to give a short overview of the signal processing implementation and the digital signal processing hardware of a hybrid system. In general, the digital hardware of a hybrid system is not different than in other wireless communication systems. Also, many steps of the signal processing are the same, e.g., the modulation, the time/frequency synchronization, and the coding/decoding. The main unique parts of the digital processing of a hybrid system are the BA and the hybrid MU-MIMO precoding/combining. Digital signal processing can be implemented in various types of digital hardware. The main types are field-programmable gate arrays (FPGAs), application-specific integrated circuits (ASICs), and microprocessors. An FPGA is a programmable digital hardware device with low latency and high signal processing performance. An ASIC is a digital IC designed for a special task and is not programmable. Due to the custom design, it has optimal performance and latency but is expensive. A microprocessor is programmable, very flexible, and easy to program. It has the highest latency which can be critical for real-time applications. Most commonly, it is used in a general form in computers.

The performance and latency requirements of the signal processing depend on the data size (e.g., the channel dimension) and the coherence period of the data (i.e., the channel). The BA and the hybrid precoding have different real-time requirements. The timing requirements of the BA depend on the coherence period of the second-order statistics of the channel $\mathbf{H}$. The coherence period of the AoA/AoD depends on the spatial consistency of the channel and the movement speed of the user. The spatial consistency of a mm-wave channel for an urban environment given as a distance is in the order of $\sim$10 m [65], [66]. Depending on the speed of the user, the coherence period of the channel second-order statistics can be between 100 ms and a couple of seconds. Thus, the BA processing has a very low latency requirement (i.e., it can be slow) and can run on all types of digital hardware, including a computer. The only important aspect is the total required training time. The longer the training time, the higher is the training overhead and the smaller is

the throughput performance of the system. The total training time is determined by the number of slots $T_{\mathrm{BA}}$ and the time per slot. The time per slot is influenced by the setup time of the beamforming and the measurement time. The measurement time is usually fixed by the length of the measurement sequence (e.g., the length of the pseudo-noise sequence in Section 3.2.1) plus some processing overhead. The setup time of the beamforming depends on the interface between the processor controlling the measurement and the analog network. The setup time of the beamforming and the measurement overhead should be optimized. For example, the measurement could be controlled by a low latency FPGA design, whereas the Non-Negative-Least-Squares problem could be solved on a computer.

The hybrid precoding matrix $\mathbf{W}$ is based on the AoD and the instantaneous effective channel $\tilde{\mathbf{H}}$. The coherence period of the instantaneous channel can be determined from the speed of the user and the center frequency. For a moderate speed of $50\,\mathrm{km\,h^{-1}}$ and a center frequency of 27 GHz, the approximate coherence period is $T_{\mathrm{coh}} \approx 400\,\mu\mathrm{s}$.[20] The coherence period of the effective channel is much shorter than the coherence time of the second-order statistics of the full channel. The beamforming part $\mathbf{U}_{\mathrm{RF}}$ of the hybrid precoding, which is based on the AoD, has a low latency requirement. The analog beamforming control can be run on a computer. The digital precoding part $\mathbf{W}_{\mathrm{BB}}$, which is based on $\tilde{\mathbf{H}}$, has to be updated for each coherence period $T_{\mathrm{coh}}$. The latency is critical and the required processing power is not negligible. Hence, it runs typically on an FPGA or an ASIC. A detailed analysis of the complexity and latency of the digital zero-forcing precoder was published in [67]. An advantage of the HDA approach is, that the effective channel $\tilde{\mathbf{H}}$ has a smaller dimension and the processing requires less computational power to run.

## 3.5. Summary

In this chapter, we have discussed multiple aspects of hybrid systems. We have described algorithms for the initial channel acquisition phase and the data communication phase. We analyzed the performance of these algorithms and some of their parameters. The selected BA algorithm achieves very good performance even for low SNR values. We simulated two variants of the hybrid data precoding algorithm. The BZF variant based on zero-forcing in the digital part achieves the best performance for nearly all scenarios (analog network structures and SNR regimes). Besides the signal processing, we also covered the hardware of the analog beamforming network. We described two structures of the analog network and considered different aspects of the components of the analog part. We analyzed the signal processing performance, the power efficiency, and the complexity for the two structures of

[20] $T_{\mathrm{coh}}$ depends on the maximum Doppler frequency $f_{\mathrm{D,max}} = f_{\mathrm{c}}(v_{\mathrm{max}}/c_0)$ with $T_{\mathrm{coh}} \approx 1/(2f_{\mathrm{D,max}})$.

the analog network. The OSPS structure has a similar performance as the FC structure but a much lower complexity and a better power efficiency. Based on our findings, we recommend an analog network with an OSPS structure. The analog network should be able to modulate the phase and the amplitude with a resolution of 4 bits for a system with 64 antenna elements. The digital precoding should use the BZF scheme. In the following chapters, we will use the results of this chapter to design complete hybrid systems. We will analyze their performance for two application scenarios.

# 4. A Prototype of the Analog Hardware

We developed a prototype of an analog beamforming module. This prototype is the core of a hybrid testbed to investigate different aspects of the HDA approach. The testbed consists of four parts, which are the antenna array, the analog beamforming network, the RF chains, and the host PC. The first version of the testbed was published in [35]. The testbed was used to investigate different simple BA algorithms by experiment and to measure their performance.

The antenna of the testbed is an array of microstrip patches with a tunable center frequency between 2.2 GHz and 2.5 GHz. The host PC can control the center frequency. The bandwidth around the center frequency is ∼30 MHz. The antenna can be used with both polarizations although only one can be enabled at a time. The antenna is built of panels of 2 by 8 elements with a spacing of $\lambda/2$ (at a frequency of 2.4 GHz) between the elements. Multiple panels can be arranged to form different antenna configurations like 2 by 16 or 4 by 8 elements. The antenna configuration for the BA experiments was 2 by 16. The antenna was not developed by the author of this thesis. The core of the testbed is an analog beamforming module solely developed by the author of this thesis. The host PC controls the system and can be used for the digital signal processing. The RF chains of the testbed published in [35] were implemented using a commercial software-defined radio (SDR). Figure 4.1 shows a picture of the testbed during the measurements for [35]. The antenna and the analog beamforming module are also part of a system that was presented at the demo session of MobiCom 2020 [38]. This system and the overall testbed are further investigated in Chapter 6.

In the following chapter, we describe the hardware design of the analog beamforming module. The focus of the thesis is not hardware design but the system concept of the HDA architecture. We describe the hardware to give an example and to introduce its parameters. The module has to be seen as a part of the whole system and was developed for experimentation. By itself, it is not comparable to research designs published in the integrated circuit/hardware scientific community.

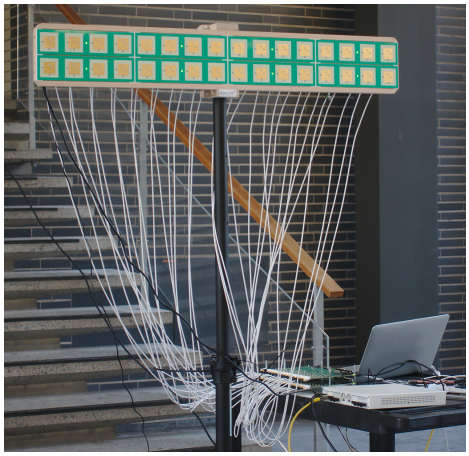


Figure 4.1.: A picture of the hybrid testbed with 32 antenna elements.

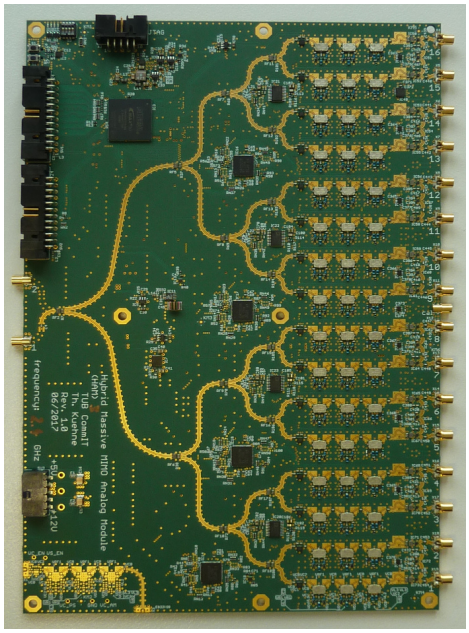



Figure 4.2.: A picture of the analog module prototype.

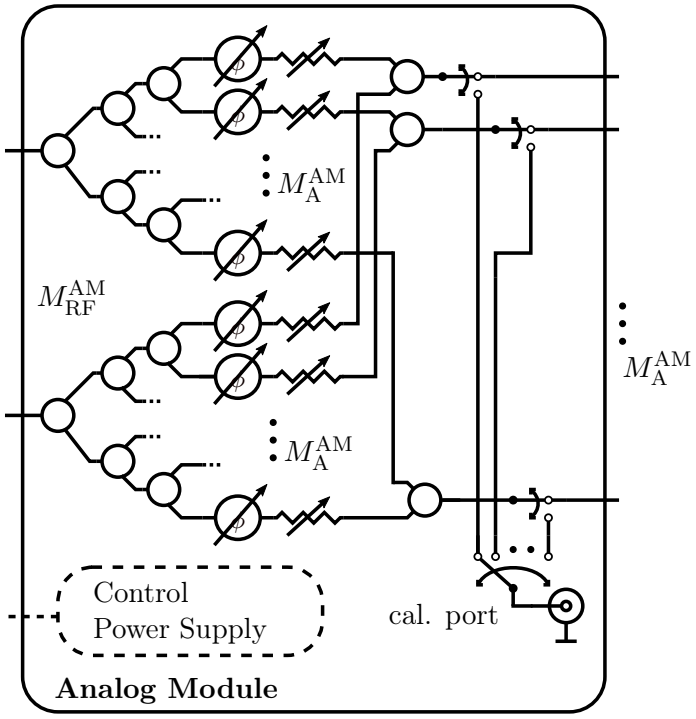



Figure 4.3.: Block diagram of the analog module prototype.

## 4.1. Hardware Design

The analog network of the prototype is based on a self-developed analog beamforming module which we will call analog module (AM). Figure 4.2 shows a picture of an AM. The analog network of the AM has a fixed size. Larger networks, e.g., with more antenna elements, can be created connecting multiple modules. Such a modularization allows for scalability of the system, different array arrangements, and ease of development of the AM. An AM is a fully connected analog RF network between $M_{\mathrm{RF}}^{\mathrm{AM}} = 2$ ports connected to RF chains and $M_{\mathrm{A}}^{\mathrm{AM}} = 16$ ports connected to antenna elements. A junction in this network is realized with a power divider/combiner. The AM has a fully connected structure. Each path of the network has a variable phase and amplitude. A simplified block diagram of the AM can be seen in Figure 4.3. To implement $M$ antennas and $M_{\mathrm{RF}}$ RF chains connected in an FC structure a system requires $M_{\mathrm{AM}} = M_{\mathrm{RF}}/M_{\mathrm{RF}}^{\mathrm{AM}} \cdot M/M_{\mathrm{A}}^{\mathrm{AM}}$ analog modules. For example, for 32 antennas and 4 RF chains connected in an FC structure in total $M_{AM} = 4$ modules are used. The module would be connected by external power dividers/combiners. The OSPS structure can be implemented if one of the two RF chain ports is left free.[1] The OSPS structure would use $M_{\mathrm{AM}'} = M/M_{\mathrm{A}}^{\mathrm{AM}}$ AMs. A mixed structure, in between the FC

[1]This is the case due to the fixed FC structure of a single module.

and the OSPS structure, can also be implemented and is for example used in the hybrid system which will be introduced in Chapter 6.

The network of a single AM has one vector modulator per path. The network and the vector modulators are reciprocal and thus the AM can be used for TDD systems without the need for switches. The distribution/combination network is realized using Wilkinson power dividers. The phase shifter design is based on a typical varactor diode architecture. A 90°-hybrid coupler divides the input signal into two branches with a phase difference of 90°. The two branches are terminated with varactor diodes. The varactor diodes reflect the signals with a certain phase change. The phase change is dependent on a control voltage applied to the varactor diodes. The reflected signals are in phase combined by the 90°-hybrid coupler at the output. At the input, the reflected signals are 180° out of phase and cancel out. The phase change between the input and output signal is equivalent to the phase change of the reflected signals by the varactor diodes. An inductor in series with each varactor diode increases the tunable phase range. One phase shifter has a tuning range between 190° and 210°. Therefore in the AM two of these phase shifters per path are used in series for a tuning range above 360°. The topology of the variable attenuator is very similar to the phase shifter. It uses PIN diodes and a 90°-hybrid coupler. Instead of the varactor diodes, the PIN diodes change not the phase but the amplitudes of the reflected signals. This effect attenuates the output signal but does not reflect the signal to the input. The power is absorbed by the PIN diodes. Figure 4.4 shows the annotated schematic of a vector modulator.

The AM initially supports a frequency range of 2.32 GHz to 2.48 GHz. Changing certain components of the AM (as can be seen in Figure 4.4) can change the frequency range. This allows manufacturing modules for different frequency bands. Additional to the 2.4 GHz version, we have planned and simulated a version for a frequency band 3.4 GHz to 3.8 GHz. The bandwidth of the AM is depending on the tolerable phase and amplitude errors. The following error and resolution values are valid for a bandwidth of 20 MHz.

The phase range of the AM is 360° and the attenuation range is 20 dB. The design of the phase and attenuation control results in a resolution of more than 9 bits for the phase and the amplitude. The practical operation resolution is lower and restricted by the phase and amplitude errors. These errors occur not only within the control range but also between the vector modulators of all paths. The module has to be initially calibrated to compensate for variations in component values, length differences, and other hardware imperfections. The calibration algorithm is explained in the following section. After calibration, the AM has a resolution of 8 bits. The AM can also be calibrated to work only as a network using phase shifters. In this mode, no amplitude control is available. On the other hand, the phase resolution does increase to 9 bits.

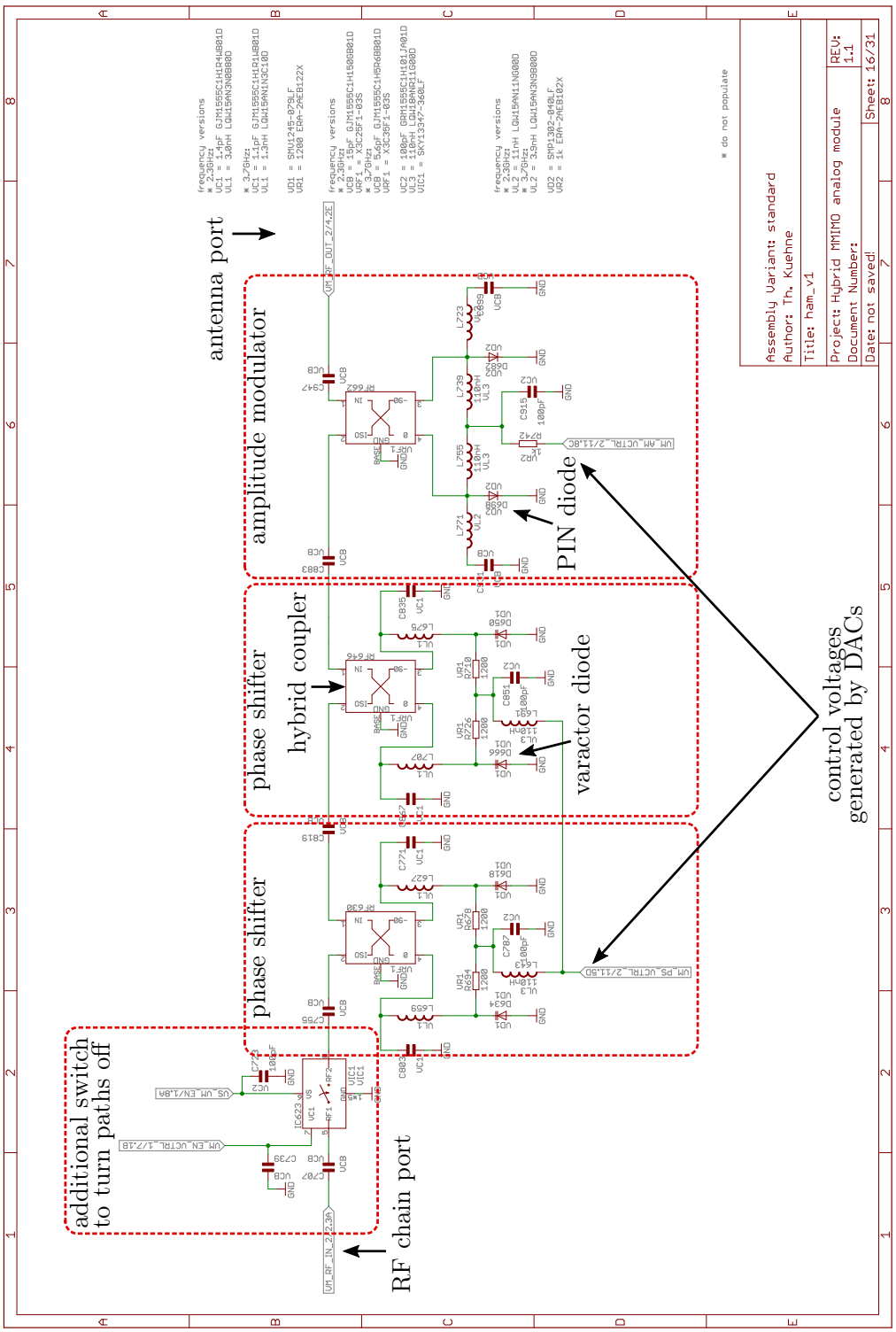



Figure 4.4.: Schematic of a single path in the analog network of the AM prototype.

The insertion loss of a calibrated AM measured between two ports is around 28 dB. The exact value does vary with the calibration but is known after a calibration run. The insertion loss has two main parts, the power distribution network and the insertion loss of the components. The loss per path of the ideal power distribution network is 15 dB. 12 dB due to the one RF chain to 16 antennas network and 3 dB due to the two RF chains per antenna network. Among others, two of the reasons for the high additional insertion loss of 13 dB are the physical size and the network size of the AM. The phase shifter and attenuator have a combined insertion loss between 5 dB and 8 dB depending on the phase setting. Furthermore, the connectors, calibration switches, imperfections of the power distribution, and the transmission lines (≈200 mm per path) add 5 dB. The AM is a research platform and not optimized in terms of loss. The high insertion loss could be nullified using a higher output power at the RF chains.

Between the vector modulators and the output ports, a calibration network was added. Through a switching matrix, the 32 paths can be connected to a single calibration port. This enables to run a calibration of all vector modulators without the need to connect all 16 antenna ports to a calibration device. This design also enabled a calibration scheme that can run during the normal operation of the hybrid system.

The module has a digital interface to control the phase and amplitude of every vector modulator. A small FPGA in conjunction with DACs converts this interface to analog control signals for the phase shifters and variable attenuators. Through the FPGA design, the interface can be adapted to the connected control instance, e.g., am SDR or a host PC. The FPGA incorporates memory and logic for storing and using the calibration data.

Table 4.1 summarizes the key parameters of the AM. The complete schematic of the AM is given in Appendix A.

## 4.2. Calibration

As already written in Section 3.3.1, most analog beamforming networks require calibration. A calibrated vector modulator has a fixed mapping between a control word (or two, for phase and amplitude) and the realized phase and amplitude changes. Calibration is required if the mapping is unknown. The exact mapping for the AM described in this chapter is unknown. The phase and the amplitude are controlled by analog voltages which are generated by DACs. The phase DACs have a resolution of 14 bits and the amplitude DACs a resolution of 16 bits (for details, see the full schematic in Appendix A). A general mapping could be derived from a circuit simulation of the vector modulators. Unfortunately, the non-ideal behavior of the components (e.g., the PIN diodes, the varactor diodes, and the

Table 4.1.: Analog Module Parameters.

| Parameter | Value |
|---|---|
| RF chain ports | 2 |
| antenna ports | 16 |
| frequency range (2.4 GHz version) | 2.32 GHz to 2.48 GHz |
| bandwidth (to guarantee the below errors) | 20 MHz |
| phase range | 360° |
| typical attenuation range | 20 dB |
| phase and attenuation resolution | 8 bits |
| phase RMS error | $< 1.4°$ |
| attenuation RMS error | $< 0.12$ dB |
| insertion loss (including 15 dB power distribution) | 28 dB |

hybrid couplers) compared to the ideal simulation falsifies the mapping. A calibration procedure can be used to measure the mapping for an individual vector modulator.

Even with a known mapping, in a real analog beamforming network, the realized phase and amplitude changes deviate from the selected phase and amplitude. We call this deviation phase and amplitude errors. The phase and amplitude errors result in malformed beam patterns. The impact of these errors is similar to the impact of the phase and amplitude quantization as considered in Section 3.3.1. The phase and amplitude errors arise from component tolerances, the non-ideal behavior of the components (e.g., over the bandwidth), differences between the vector modulators in the analog network (e.g., due to different transmission line lengths), and other sources. A calibration procedure can be used to measure and correct these errors up to residual phase and amplitude errors. The size of the residual phase and amplitude errors depend on the measurement accuracy of the calibration equipment, the impact of thermal noise in the circuit, and non-ideal behavior which can not be corrected. In a well-calibrated system, the residual phase and amplitude errors are smaller than the phase and amplitude resolutions.

In the following, we describe the calibration procedure used to measure the mapping of the control words to the phase/amplitude values for the AM. Each physical AM is calibrated on its own. Since the AM is physically reciprocal, the calibration only measures one direction. The AM is connected to a two-port vector network analyzer (VNA) as can be seen in Figure 4.5. One VNA port is connected to the calibration port of the AM. The other VNA port is connected to one of the two RF chain ports of the AM. During the calibration measurement, the RF cable is switched between the two RF chain ports. The

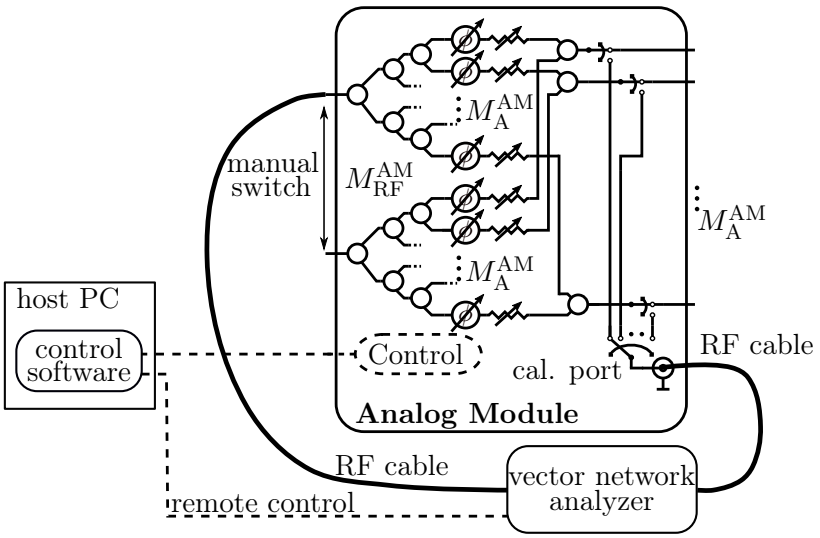



Figure 4.5.: Block diagram of the calibration setup of the AM.

**Algorithm 1:** The measurement procedure for the AM calibration.

**Input:** measurement and system parameters (frequency, $M$, $M_{\mathrm{RF}}$,...)
**Input:** amplitude ($\mathcal{A} \in \mathbb{Z}^{N_{\mathrm{a}}}$) and phase ($\mathcal{P} \in \mathbb{Z}^{N_\phi}$) measurement codebooks of control words

**Initialize:** setup the AM and the VNA
**Initialize:** disable all vector modulators

**for** $r = 1, 2, \ldots, M_{\mathrm{RF}}$ **do** ▷ *loop over all RF chain ports*
  connect RF chain port $r$ to the VNA
  **for** $m = 1, 2, \ldots, M$ **do** ▷ *loop over all antenna ports*
    enable vector modulator $(r, m)$
    connect vector modulator $(r, m)$ to the calibration port
    **for** $a = 1, 2, \ldots, N_{\mathrm{a}}$ **do** ▷ *loop over* $\mathcal{A}$
      ▷ *Set amplitude control of the vector modulator* $(r, m)$ *to the* $a$*-th value.*
      $A_{\mathrm{VM},r,m} = [\mathcal{A}]_a$
      **for** $p = 1, 2, \ldots, N_\phi$ **do** ▷ *loop over* $\mathcal{P}$
        ▷ *Set phase control of the vector modulator* $(r, m)$ *to the* $p$*-th value.*
        $\phi_{\mathrm{VM},r,m} = [\mathcal{P}]_p$
        $[\mathbf{D}_{r,m}]_{a,p} = \mathrm{VNA}$ ▷ *Measure the complex coefficient with the VNA*
      **end**
    **end**
    disable vector modulator $(r, m)$
  **end**
**end**

**Output:** measurement data $\mathbf{D}_{r,m} \in \mathbb{C}^{N_{\mathrm{a}} \times N_\phi}$ for all $r \in [M_{\mathrm{RF}}]$ and $m \in [M]$ vector modulators.

VNA and the AM are remotely controlled from the host PC. The calibration is carried out at a specific frequency. The measurement procedure is described in Algorithm 1. The vector modulators of an AM are measured successively. The complex transmission coefficients are measured with the VNA for a two-dimensional sweep through the phase and amplitude control ranges. The phase range is between 0 and 15 496 and the amplitude range is between 0 and 38 010.[2] Per sweep, 256 values are measured (16 phase and 16 amplitude values). The size of the phase and amplitude sweep is limited due to the measurement time. The time duration of each measurement depends on the VNA settings, the VNA interface time, and the AM control time. For the following calibration results, the total measurement time of a single AM is around one hour.

The measurement data is used to derive the control words for a given phase resolution, amplitude resolution, and amplitude/attenuation dynamic range. Algorithm 2 describes the calculation procedure. All vector modulators of an AM are collectively calibrated and common phase and amplitude offsets are calculated. Due to the restricted size of the phase and amplitude measurement sweep, the measurement data is first interpolated to the full control ranges. Afterward, the common phase and amplitude offsets are calculated for all vector modulators over the full phase and amplitude control ranges. The calibration procedure computes the error distance between the interpolated measurement data and the phase and amplitude target pair. The control words (phase and amplitude) of the minimum error distance are stored for all target value pairs.

The frequency dependence of the vector modulator is a non-ideal behavior that can not be corrected by the calibration scheme. The AM does not use true time delay phase shifters but phase shifters with a narrowband assumption (see Section 3.3.1). The phase and amplitude errors increase as the difference between the signal frequency and the calibration frequency grows. The achieved resolution and the residual errors are valid within a specific bandwidth (i.e., 20 MHz). For the whole frequency range of operation, multiple frequency points need to be calibrated.

The AM, calibrated with the described procedure, achieves a resolution of 8 bits for the phase and the amplitude with an amplitude/attenuation range of 20 dB. After running the calibration procedure, we measured the residual phase and amplitude errors between the RF chain ports and the antenna ports for all vector modulators of an AM. Figure 4.6 shows the phase error and the amplitude error for one exemplary vector modulator at a frequency of 2.4 GHz over the full settings range. The performance of a complete AM over the 20 MHz bandwidth can be seen in Figure 4.7. Figure 4.7 shows the maximum and the root mean square (RMS) of the errors of all vector modulators and all phase/amplitude

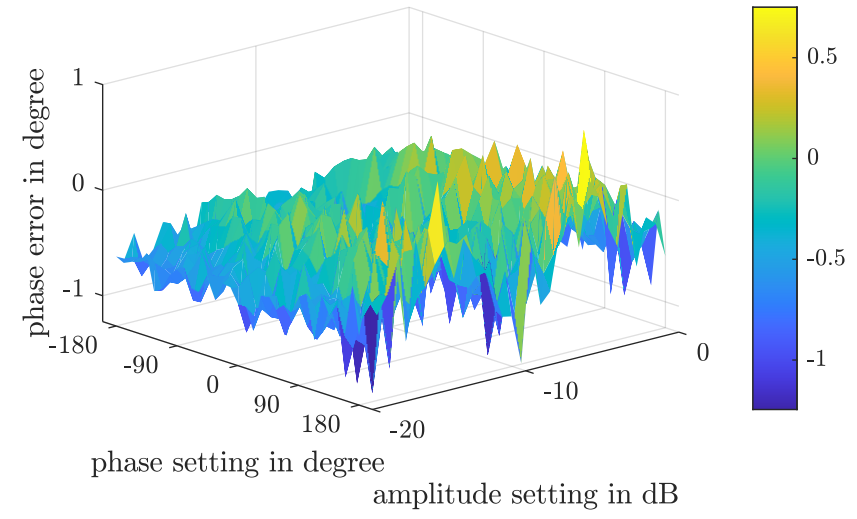
[2]The ranges depend on the maximum voltages of the diodes, the DAC reference voltages, and the DAC resolutions.

**Algorithm 2:** The calculation procedure for the AM calibration.

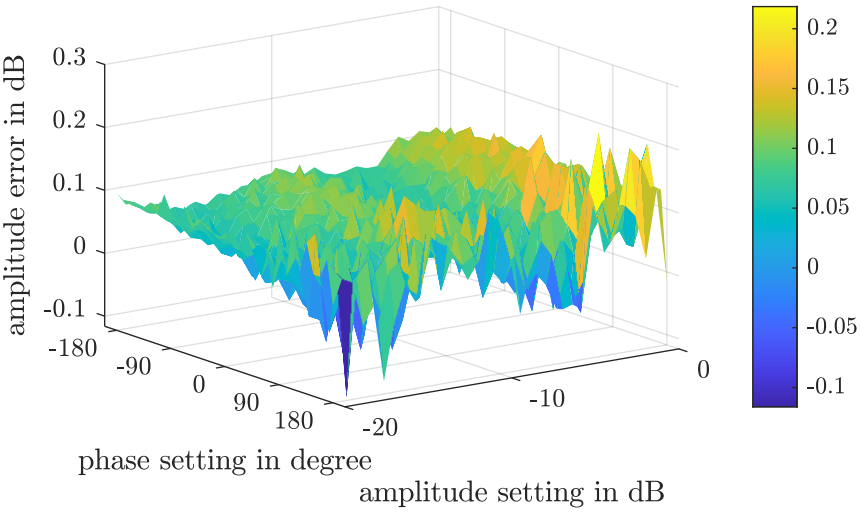
**Input:** $\mathbf{D}_{r,m}$ for all $r \in [M_{\mathrm{RF}}]$ and $m \in [M]$ vector modulators
**Input:** amplitude ($\mathcal{A} \in \mathbb{Z}^{N_{\mathrm{a}}}$) and phase ($\mathcal{P} \in \mathbb{Z}^{N_\phi}$) measurement codebooks
**Input:** target amplitude ($b_a$) and phase ($b_\phi$) bit width
**Input:** $A_{\max}$ amplitude dynamic range in dB

**for** $r = 1, 2, \ldots, M_{\mathrm{RF}}$ **do** ▷ *loop over all RF chain ports*
  **for** $m = 1, 2, \ldots, M$ **do** ▷ *loop over all antenna ports*
    $\mathcal{A}_{\mathrm{int}} = \{0, 1, \ldots, \max(\mathcal{A})\}$
    $\mathcal{P}_{\mathrm{int}} = \{0, 1, \ldots, \max(\mathcal{P})\}$
    ▷ *cubic spline interpolation of* $\mathbf{D}$ *from* $\mathcal{A}$ *and* $\mathcal{P}$ *sample codewords to all possible codewords*
    $\mathbf{D}^{\mathrm{int}}_{r,m} = \texttt{interpolate}(\mathcal{A},\mathcal{P},\mathbf{D}_{r,m},\mathcal{A}_{\mathrm{int}},\mathcal{P}_{\mathrm{int}})$
  **end**
**end**
▷ *common phase and amplitude offsets of all vector modulators*
$\phi_{\mathrm{off}} = -1 \cdot \max\limits_{r,m,\mathcal{A}_{\mathrm{int}}} \left( \min\limits_{\mathcal{P}_{\mathrm{int}}} \left( \arg(\mathbf{D}^{\mathrm{int}}) \right) \right)$
$A_{\mathrm{off}} = 1 / \left( \min\limits_{r,m,\mathcal{P}_{\mathrm{int}}} \left( \max\limits_{\mathcal{A}_{\mathrm{int}}} \left( |\mathbf{D}^{\mathrm{int}}| \right) \right) \right)$
**for** $r = 1, 2, \ldots, M_{\mathrm{RF}}$ **do** ▷ *loop over all RF chain ports*
  **for** $m = 1, 2, \ldots, M$ **do** ▷ *loop over all antenna ports*
    **for** $a = 1, 2, \ldots, 2^{b_{\mathrm{a}}}$ **do** ▷ *loop over the target amplitude values*
      **for** $p = 1, 2, \ldots, 2^{b_\phi}$ **do** ▷ *loop over the target phase values*
        $\phi_{\mathrm{t}} = \frac{2\pi\,(p-1)}{2^{b_\phi}-1}$ ▷ *target phase*
        $A_{\mathrm{t}} = 10^{\frac{A_{\max}\,(a-1)}{2^{b_{\mathrm{a}}}-1}/20}$ ▷ *target amplitude*
        ▷ *find the best fitting control word for the target*
        $x, y = \arg\min\limits_{\mathcal{A}_{\mathrm{int}},\mathcal{P}_{\mathrm{int}}} \left( |\mathbf{D}^{\mathrm{int}}_{r,m} \cdot A_{\mathrm{off}}\, e^{i\phi_{\mathrm{off}}} - A_{\mathrm{t}}\, e^{i\phi_{\mathrm{t}}}| \right)$
        $[\mathcal{A}^{\mathrm{ctrl}}_{r,m}]_{a,p} = x$
        $[\mathcal{P}^{\mathrm{ctrl}}_{r,m}]_{a,p} = y$
      **end**
    **end**
  **end**
**end**

**Output:** $\mathcal{A}^{\mathrm{ctrl}}_{r,m} \in \mathbb{Z}^{2^{b_a} \times 2^{b_\phi}}$ amplitude and $\mathcal{P}^{\mathrm{ctrl}}_{r,m} \in \mathbb{Z}^{2^{b_a} \times 2^{b_\phi}}$ phase control words for all $r \in [M_{\mathrm{RF}}]$ and $m \in [M]$ vector modulators.




(a)




(b)

Figure 4.6.: Measured error performance of one vector modulator of the calibrated AM with an 8-bit resolution at 2.4 GHz: (a) phase error over the settings range, and (b) amplitude error over the settings range.

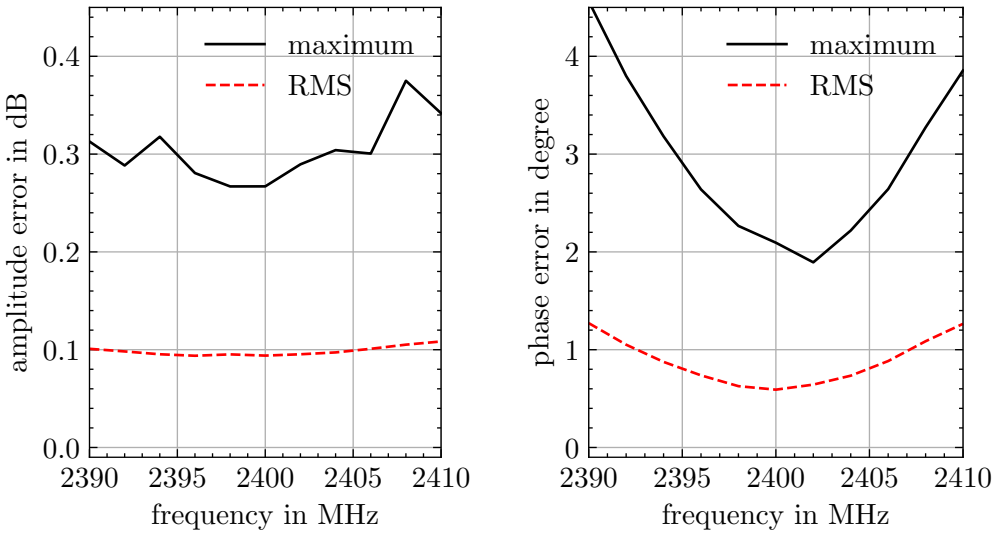



Figure 4.7.: Measured maximum and RMS of the phase error and the amplitude error over the bandwidth of an AM with 8-bit resolution calibrated at 2.4 GHz: (a) amplitude error and (b) phase error.

settings. The RMS errors are smaller than the least significant bit for an 8-bit resolution. As expected, the phase error in Figure 4.7(b) does increase for frequencies further away from the calibration frequency. Reducing the phase error even further below 1° is very challenging as the whole calibration setup has an impact on the result. This includes the measurement accuracy of the VNA and inaccuracies like a phase drift over the measurement duration. The 8-bit resolution is already higher than offered by off-the-shelf variable phase shifters and attenuators, especially, taking into account that it is achieved over the phase range, the amplitude range, and 32 vector modulators.

To visualize the impact of the real hardware including the phase and amplitude errors, we measured the channel gain of a beam sweep over the whole azimuth range of 180°. The channel was measured in a large empty room between the testbed with a 2 by 16 antenna array and a directive measurement antenna. This setup ensured a LOS condition. Figure 4.8 shows the comparison between the measurement and the calculated channel gain assuming ideal hardware and only the LOS path. Both curves show a close agreement for the LOS/main-lobe and no additional strong side-lobes. Thus, we can assume, that the testbed works very well and can be used for further investigations on the HDA architecture.

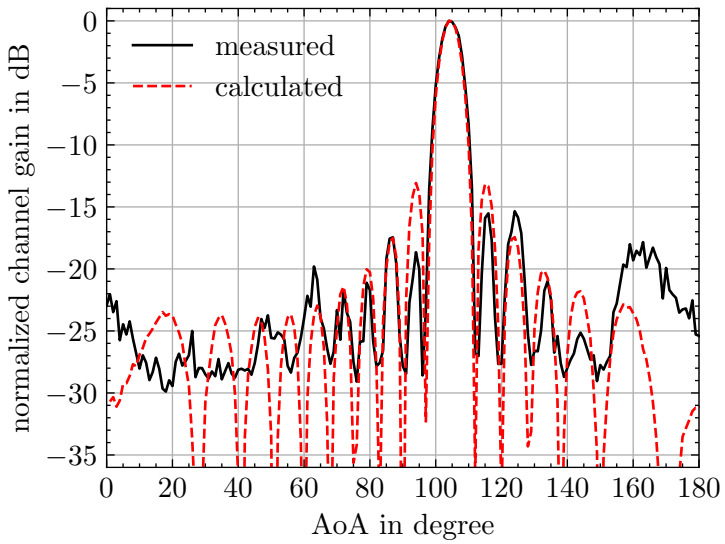



Figure 4.8.: Comparison of the measured and calculated channel gain vs. the beam sweep angle using the hybrid testbed at 2.4 GHz with a 2 by 16 antenna array configuration. The channel was measured in a LOS scenario and the calculated channel assumes a single path.

# 5. 5G mm-wave Application Scenario

The main application of the HDA architecture considered in the literature is in systems operating in the mm-wave frequency range [14]. The increased available bandwidth in the mm-wave frequency range is one of the main contributions to the increased data rate of the 5G standard [60]. 5G brought mm-wave systems for mobile communication closer to commercialization than ever. The increased bandwidth compared to the sub-6 GHz frequency range makes the fully-digital architecture infeasible for mm-wave systems with the current technology. The use of large antenna arrays and the HDA approach makes 5G mm-wave systems possible [14]. The mm-wave channel has distinct features compared to the frequency range below 6 GHz [56]. It has a considerably larger free-space path loss due to the high frequency. In conjunction with the high path loss, the channel above the noise floor consists of fewer paths from fewer scatterers and the impact from blockage is much higher [56]. State-of-the-art systems use smaller cells and large antenna arrays to compensate for these effects and increase the system sum-rate [68]. The following analysis will focus on the scenario of an urban small cell with a diameter of up to 300 m and slowly moving users.

In recent years multiple testbeds and prototypes of mm-wave systems have been published demonstrating certain features and properties [34]. The European research project SERENA (`http://www.serena-h2020.eu/`) developed a hardware and signal processing platform for the use in mm-wave mobile communication. The goal of the project was to have a proof-of-concept for a 5G system in the 37 GHz to 40 GHz band based on an innovative hardware integration platform. In this chapterr, we evaluate using simulations the hypothetical performance of a system based on the SERENA integration platform. Compared to more theoretical investigations in Chapter 3, we incorporate additional realistic assumptions on hardware impairments like non-ideal antenna patterns and design parameters like the array size. Furthermore, where possible we use physical layer parameters defined by the 5G standard. We use the QuaDriGa 3-D channel simulator [69] with 3GPP parameters for the mm-wave channel. This chapter is based on the publication “Performance Simulation of a 5G Hybrid Beamforming Millimeter-Wave System” [20].

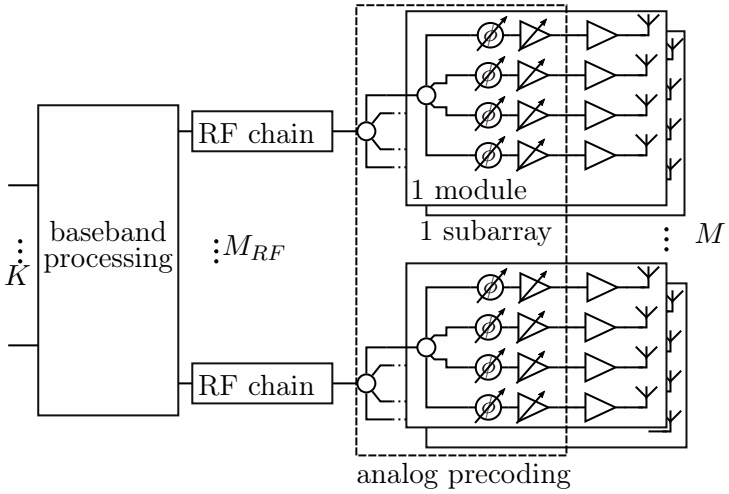



Figure 5.1.: Block diagram of the OSPS transmitter architecture of the SERENA system.

## 5.1. System and Hardware Description

As mentioned above, the described system is part of the research project SERENA. Hence, the simulation and the system design in this paper follow closely the hardware specifications of the project [37], [70]. The hardware design is described below. The signal processing as described in Section 3.2 is adopted to conform with the 5G physical layer specifications. Figure 5.1 shows the structure of the system. It is an implementation of the OSPS structure for the analog network as introduced in Section 3.1. The hardware partners in the SERENA project declared the OSPS structure the only feasible option. As the theoretical investigation in Chapter 3 showed, the performance loss compared to the FC structure is negligible.

The core part of the hardware design is a fully integrated module in the form of a small printed circuit board. It has one RF port connected to four antenna elements by phase shifters and amplitude modulators. The antennas are on top of the module. The module is designed for TDD operation in the 5G frequency band between 37 GHz and 40 GHz. The transmitter part is shown as a block diagram in Figure 5.1. The receiver is not shown since we focus on the downlink. Transmitter/receiver reciprocity, which is required by parts of the signal processing, requires a calibration procedure is not described here. It is very similar to the procedure described in Section 4.2. The hardware simulation is based on the design specifications [37].

The module incorporates a beamforming chip with four channels. It can optionally have four frontend chips with a power amplifier and a low noise amplifier. The power amplifiers boost the maximum output power per beamforming channel to above 33 dBm. This could enable a large BS cell coverage, possibly larger than common small cells. Since

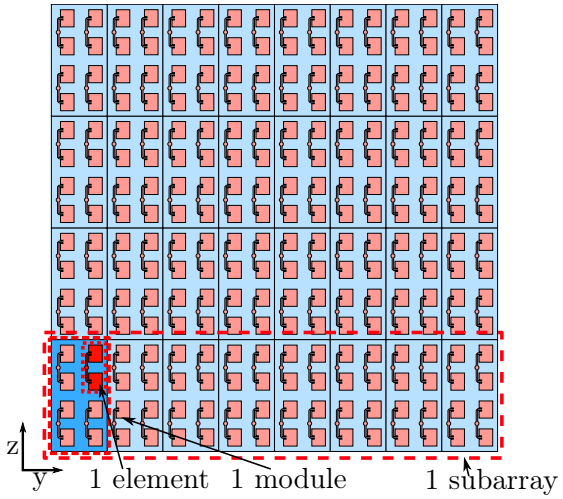
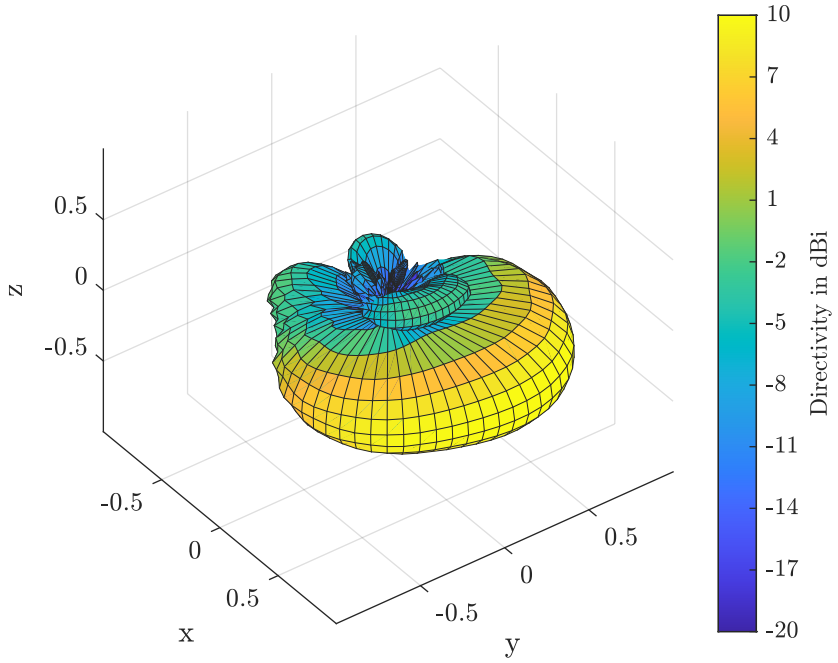



Figure 5.2.: Array configuration of the BS showing the module arrangement.

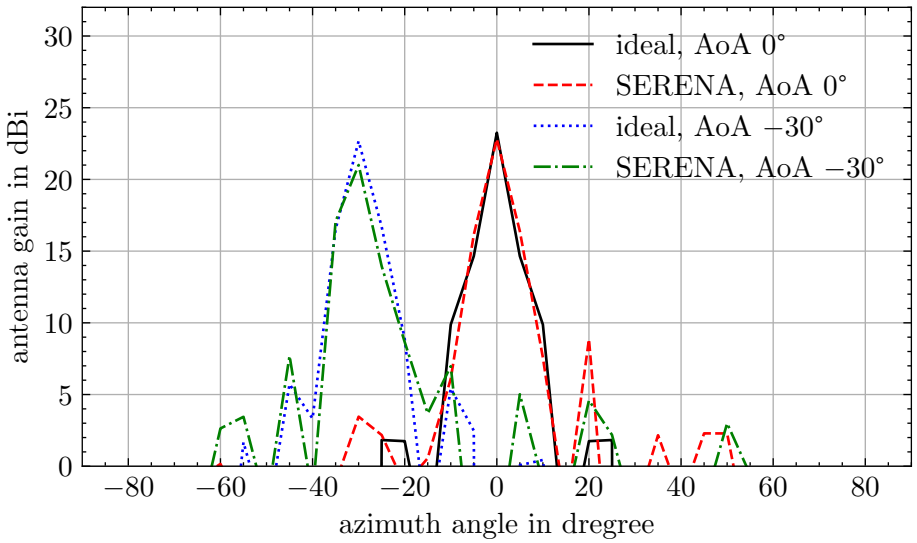
in this chapter we focus on the small cell scenario, we simulate the modules without power amplifiers. The maximum output power of the beamforming chip is 10 dBm. In our simulations we include a power backoff of 12 dB for a typical 5G waveform [71] resulting in a maximum output power of a single channel of $P_{\mathrm{o,amp}} = -2\,\mathrm{dBm}$. The beamforming capabilities of the module include phase and amplitude modulation. The phase resolution is 5° within a range of 0° to 360°. The amplitude can be set with 0.5 dB steps in a 30 dB range.

Each of the four antenna elements in the module consists of two microstrip patches with horizontal polarization and a half-wavelength spacing. The two patches are fed with equal phase and amplitude from a single beamforming channel using a power divider. Thus, they can be seen as a single element with a specific (directional) radiation pattern (see Figure 5.3(a)). The antenna was designed at the Fraunhofer Institute for Reliability and Microintegration (IZM). The overall size of the module is half of the wavelength at the center frequency times the number of patch elements in the respective direction. Hence, multiple of the modules can be used together to create larger antenna arrays. Figure 5.2 shows the concept and the size simulated in this paper. The module concept simplifies the manufacturing and enhances the RF performance but also gives flexibility in designing a larger array.

In the simulations we investigate an array (Figure 5.2) which consists of four subarrays connected to $M_{\mathrm{RF}} = 4$ RF chains. Each subarray is built with eight modules resulting in $\hat{M} = 16$ horizontal elements and two vertical elements. During the simulation we only consider the horizontal domain and use the vertical elements of a subarray with equal




(a)




(b)

Figure 5.3.: Antenna patterns of the SERENA system: (a) 3-D view of a single element pattern and (b) example beam patterns of a single subarray for the SERENA array and an array of ideal patch elements.

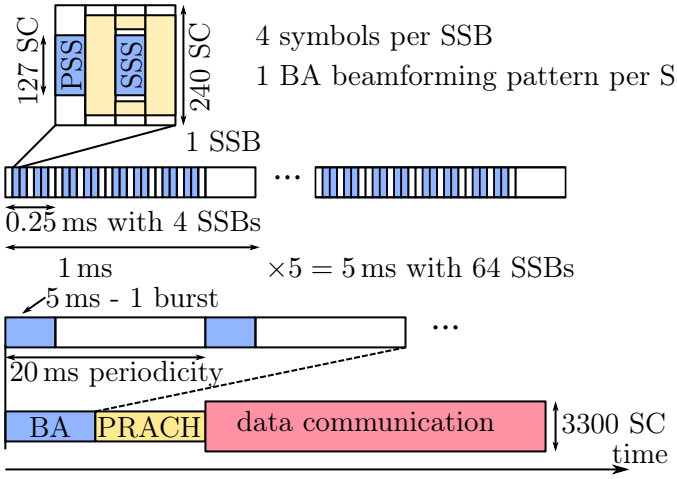



Figure 5.4.: Frame structure of the signal processing adopted to the 5G standard for 120 kHz SC spacing.

phase and amplitude resulting in a fixed beam in elevation in the broadside direction. The total number of steerable elements in the array is $M = \hat{M} \cdot M_{\mathrm{RF}}$ (considering two vertical elements in Figure 5.2 as one).

## 5.2. Signal Processing and Frame Structure

A critical part of a 5G mm-wave system is the signal processing necessary for MU-MIMO, this includes the measurement of the channel state and the precoding. As mentioned before, due to the large path loss and the HDA structure of the system the channel is estimated using a BA algorithm. The result of the BA is a pair of beam directions connecting the BS and each UE. We name the direction at the BS as AoD and at the UE as AoA. In the data communication phase the BS and the UE use beamforming towards these directions to increase the SNR and overcome the large path loss. In this section, we use the BA algorithm described in Section 3.2.1. The data communication is based on the hybrid precoding investigated in Section 3.2.2.

The time flow of the whole communication with the different phases is set by the protocol and the frame structure. Figure 5.4 shows the frame structure envisioned for the proposed system. It is based on the 5G definition for the initial acquisition as described in [60]. We use the configuration with the highest available bandwidth of the standard with a subcarrier (SC) spacing of 120 kHz, resulting in 400 MHz channel bandwidth. During the BA phase the synchronization signal blocks (SSBs) are used as training symbols. One SSB is equivalent to one training slot in the described BA algorithm and consists of two

synchronization sequences which can be used to estimate the channel power. The encoded information in an SSB can be used by the UE to identify the index of the current pattern from the BS pseudo random beamforming codebook. The 5G initial acquisition structure supports 64 training slots in a 5 ms SSB burst. The BS can transmit every 20 ms an SSB burst to increase the number of available training slots to more than 64. Once a UE has determined an AoD and AoA pair, it feeds the AoD information back to the BS using the physical random access channel (PRACH). The UE transmits the feedback using a single beam along the calculated AoA. During the PRACH phase the BS uses an omnidirectional pattern or sweeps through multiple beam directions to receive the UE feedback. Once the BS has received the AoD information from the UE it can schedule the user in the downlink data communication phase using MU-MIMO precoding as described below.

As can be seen in Figure 5.4, in 5G the SSBs use 240 SCs and the data uses 3300 SCs. Within the BA phase of an HDA system the BS can not combine SSB and data SCs since the analog beamforming is independent of the SCs. Thus, the power per SC of a SSB is ∼11 dB higher than of a data SC improving the measurement SNR of the BA phase. The path directions in the underlying channel are consistent with the channel second-order statistics and vary much slower than the channel small-scale fading [56]. Hence, the data communication phase can be much longer than the BA phase in order to reduce the training overhead[1].

After the successful BA, a scheduler selects a subset of users for the data communication. The directional nature of the mm-wave channel and the HDA system demand a directional scheduler which selects UEs with a required minimum angular separation [19]. We did not include a scheduling algorithm in our simulations but use sets of users with sufficient angular separation for the data communication. During the data communication phase, we use the precoding algorithm investigated in Section 3.2.2. We simulate both variants of the proposed algorithm (i.e., BST and BZF). We assume UEs with a single RF chain/data stream. Hence, for both algorithm variants the UE points a single beam towards the acquired AoA. The scheme could easily be extended to support multiple RF chains per user. The BS uses the analog beamforming of one subarray to point a single beam towards the AoD of the UE. The other subarrays point towards different UEs. Please note, that further parts of the baseband signal processing like e.g. the modulation, time synchronization, and instantaneous channel estimation are out of the scope of this chapter. We focus on the parts directly related to the HDA-system approach. The other parts are within the 5G standard relatively independent of the details of the HDA implementation.

[1] In [56] the authors measured a coherence distance of the second-order statistics of more than 10 m. Even with a high speed like 10 m/s (for a small cell scenario) the data communication could use the BA information for up to 1 s.

## 5.3. Simulation Environment

To analyze the performance of the described system, we setup a simulation environment based on the simulations of the signal processing algorithms in Chapter 3, EM field simulations of the hardware module, and the QuaDriGa 3-D channel simulator [69]. QuaDriGa is a geometry-based stochastic channel model incorporating, among others, features like UE mobility, 3GPP channel parameters and non-ideal antenna patterns.

The radiation patterns of the antenna elements are simulated using CST Microwave Studio©2017 by the Department of Microtechnology and Nanoscience of the Chalmers University of Technology. The 3-D directivity pattern of the vertical polarization for the element highlighted in Figure 5.2 is shown in Figure 5.3(a). The gain of one antenna element in broadside direction is 8.5 dBi. Two example radiation patterns of single direction Fourier based beams of one subarray can be seen in Figure 5.3(b). The non-ideal characteristics of the antenna elements create slightly larger side-lobes and smaller main-lobes compared to a subarray with ideal patch elements. The radiation characteristics of the full antenna array are created using concatenated simulation results from a single module. The simulated antenna characteristics are used at the BS in the QuaDriGa simulator. The UE antenna is set to an array of four ideal patch elements with a half-wavelength spacing.

Unless otherwise specified, we choose the following parameters for the simulations. The center frequency is set to 39 GHz. Other timing parameters are set to the respective values defined by the 5G standard as described in Section 5.2. The noise figure of the UE receiver is chosen to be 7 dB, which is consistent with the noise figure of the beamforming chip in a SERENA module. The temperature, used to determine the noise power, is 333 K. The BS is 10 m and all UEs are 1.5 m high. The BS array is tilted downwards with an angle of 11.6°. The phases and amplitudes for the beamforming are quantized as described in Section 5.1 with an additional equally distributed random error in the range of 5° and 0.5 dB to simulate hardware tolerances for the phase and amplitude, respectively. The QuaDriGa channel scenario is set to the values defined in the technical report 3GPP 38.901 [66] (but extended through the QuaDriGa setting `use_3GPP_baseline=0`). In this report, the 3GPP consortium standardized channel model parameters for mm-wave channels, which should be used for 5G testing. In our simulations, we use the urban micro parameters both for LOS and NLOS.

The users are distributed in a distance $d_{\mathrm{UE}}$ between 10 m and 300 m and in an azimuth range of $-45°$ to $45°$. The azimuth range is divided into $K$ equally sized subranges with an angular separation of $\Delta_{az} = 5°$. An equal number of users is randomly located in each subrange. The ideal scheduler picks one user of each subrange for the data rate simulation. Due to the four subarrays of the SERENA array, $K = 4$ UEs are scheduled

in one communication slot. During all simulations the UEs have a moving speed of 1 m/s, a slow walking speed. The random walking direction is in the range $-45°$ to $45°$ towards the BS. This guarantees that the users always have a LOS channel for a LOS channel simulation.

For the BA simulation, we use beamforming patterns with $\kappa_u = 2$ probed directions for the BS and a single beam for each UE. We allow a maximum number of 192 training slots. The result of the BA after the 192 slots is used as input for the precoding algorithm, even in case of an incorrect result.

We do not simulate a complete communication system with, for example, modulation, channel coding, and synchronization. The focus of the simulation is to characterize the hardware system and the HDA related algorithms described in Section 5.2. To evaluate the performance of the hardware and the precoding algorithm during the data communication phase, we calculate the achievable asymptotic ergodic spectral efficiency as introduced in Section 3.2.2. The number of SCs during the data communication is $F = 3300$. $P_{k,\omega}$ in (3.17) is in this setting the output power of a subarray over the number of SCs $F$ and equal to $\hat{M} \cdot P_{\mathrm{o,amp}}/F = -25.14\,\mathrm{dBm}$ for all users and SCs. The sum spectral efficiency for the $K$ scheduled UEs is $R_{\mathrm{sum}} = \sum_{k=1}^{K} R_k$. This metric does not include the overhead due to the BA and the instantaneous channel estimation, which is required by the ZF scheme. It is an upper bound and will be lower in a real system with modulation and channel coding.

## 5.4. Simulation Results

The cell configuration is visualized in Figure 5.5 (a). It shows the LOS power levels in the cell for a BS with a single element antenna transmitting the total output power of the full array and an ideal UE with an omnidirectional antenna pattern. One can see, that close to the BS, $d_{\mathrm{ue}} < 11\,\mathrm{m}$, the power decreases rapidly. Below this distance the UE is not covered by the main lobe of the BS array. This limit could be reduced by using vertical beamforming at the BS or changing the BS array tilt at the cost of the maximum distance coverage. In Figure 5.5 (b) the SNR before beamforming (BBF) is shown for the BA phase (SSB frames) and the data communication phase. BBF means, that the BS and the UEs use a single antenna element with the respective radiation pattern, the BS transmits with the total output power of the full array, and the UEs receive with the given noise figure. The SNR is simulated for the given SC spacing. The curves named “mean” are averaged over all users, SCs, and slots. The “min” and “max” curves are the minimum and maximum values of all simulated users and slots. The data curve is lower than the SSB curve due to the mentioned difference in the number of used SCs. The LOS channel scenario is shown in

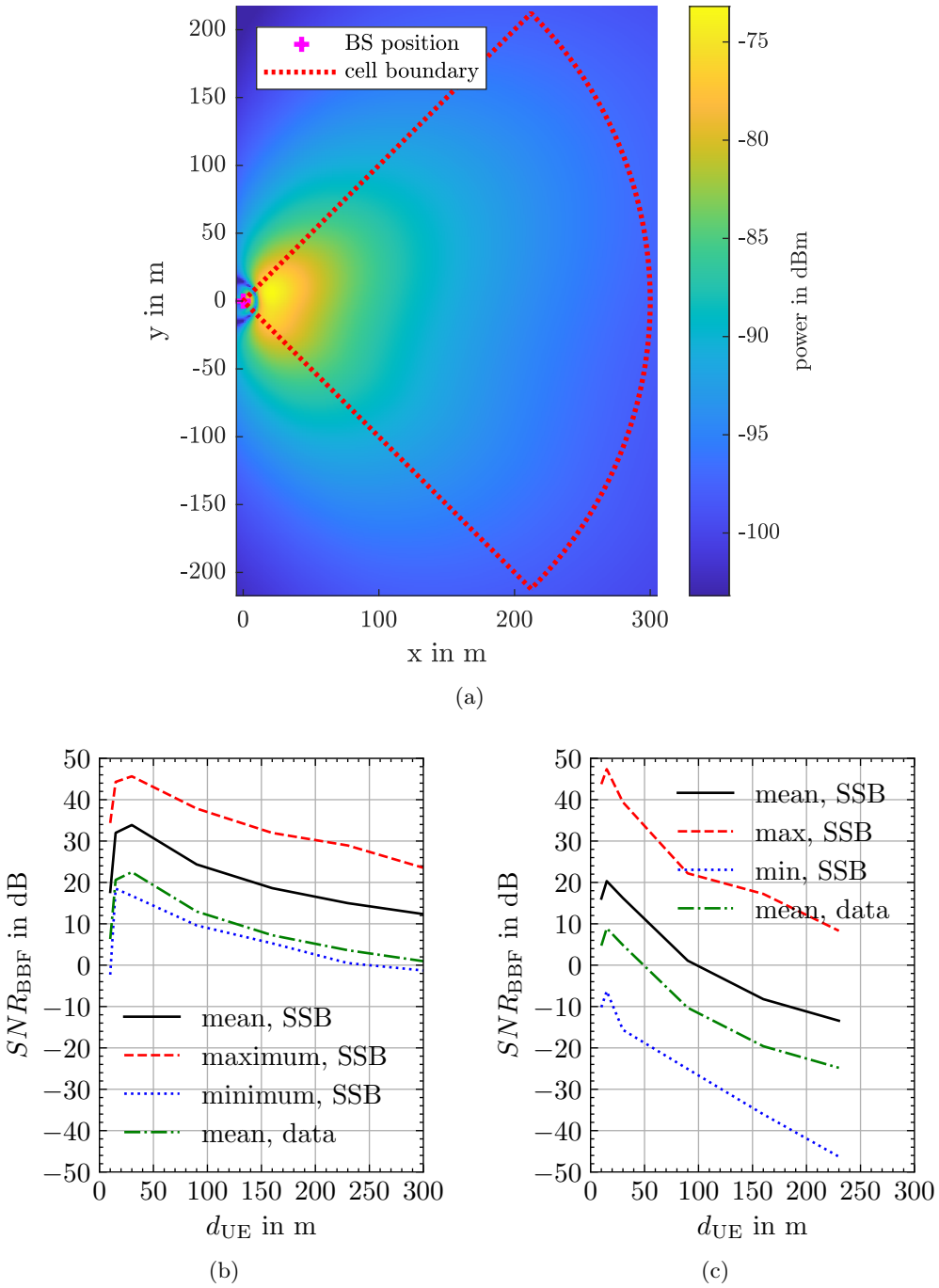



Figure 5.5.: QuaDriGa channel simulation results of the 5G system: (a) visualization of the cell, and (b) and (c) the average $\mathrm{SNR}_{\mathrm{BBF}}$ for the BA and data communication phase for the LOS and NLOS case, respectively.

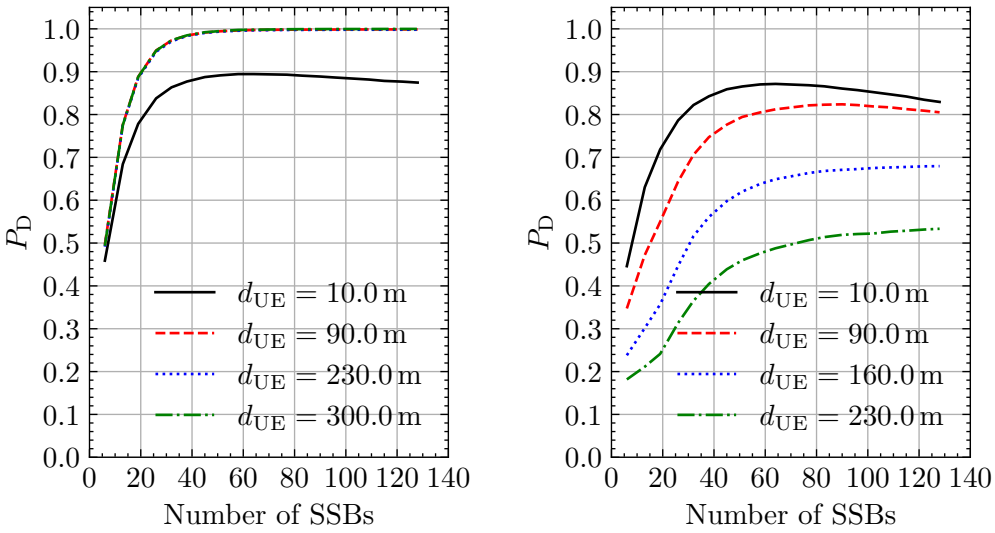



Figure 5.6.: Average detection probability $P_D$ of the BA for different user distances $d_{UE}$ for (a) the LOS case and (b) the NLOS case.

Figure 5.5(b) and the NLOS scenario in Figure 5.5(c). Naturally, the SNR of the NLOS case is much lower than for LOS.

*Beam Alignment:* The performance measure of the BA is the detection probability $P_D$. It is the probability of finding an AoD-AoA pair between the BS and the UE for which the channel loss is within a 2 dB range from the strongest pair. Multiple pairs can have a similar path loss due to the fact that the BA output is a discrete grid index while the user placement is continuous or because multiple paths have a similar loss in a NLOS channel. Since we use the simulated BA results as input of the precoding simulation this effect is included in the data rate results. The mean of $P_D$ for all UEs of a given distance $d_{UE}$ is shown in Figure 5.6. Figure 5.6(a) shows the LOS case and Figure 5.6(b) the NLOS case. In general, the simulated performance of the BA algorithm for LOS is very good. $P_D > 0.95$ is achieved for the whole cell with less than 64 slots in the LOS scenario. The bad result for LOS at $d_{UE} = 10\,m$ is due to the mentioned lower distance limit of the main lobe in elevation of the BS antenna. The main reason for the worse performance in the NLOS case is a very low SNR for some of the UEs (see the min curves in Figure 5.5). Nonetheless, the BA algorithm does estimate correct AoD-AoA pairs for near NLOS users. The small cell assumption with the LOS condition[2] together with the reduced number of SCs of the SSBs is favorable for the BA.

[2]LOS is a typical assumption for mm-wave small cell systems [68].

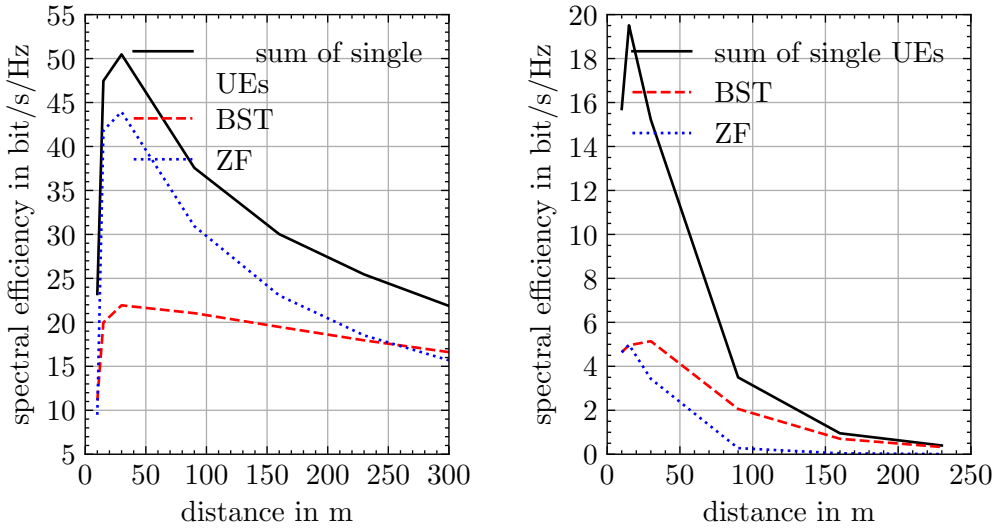



Figure 5.7.: Average sum spectral efficiency $R_{\text{sum}}$ over the UE distance $d_{\text{ue}}$ for (a) the LOS case and (b) the NLOS case.

*Data Communication:* To evaluate the data communication performance, we compare the two variants of our precoding algorithm with the sum of the single user spectral efficiency. The single user sum is the interference free upper bound for the precoding results. The average results for multiple user sets vs. the distance of the UEs to the BS is shown in Figure 5.7. In the very high SNR region (see Figure 5.5 (b), below 200 m) the ZF variant achieves a very high MU-MIMO gain[3] and has a much better performance than the BST variant. With an increasing distance, the SNR decreases and the BST achieves a similar performance. At the edge of the cell, BST shows a better performance than ZF. This result is consistent with our previous results in Section 3.2.3. Due to the reduced SNR of the NLOS scenario the achievable data rates are much lower and not the complete cell can be covered by the BS. Nevertheless, UEs closer than 150 m could be served by the BS applying the BST precoding scheme even with a very low SNR.

## 5.5. Summary

In this chapter, we presented an HDA beamforming mm-wave system which was developed in the European project SERENA. We described the system specifications, the underlying signal processing algorithms, and the frame format based on the 5G standard. We jointly

[3]Which means it is close to the sum of the single user rate.

evaluated the performance of the complete HDA system including the antenna patterns, hardware specifications, and the BA and the data precoding algorithms. We used 5G physical layer frame parameters and the QuaDriGa 3-D channel simulator with 3GPP mm-wave channel parameters. The system shows in the simulated small cell scenario a very good BA performance and a good MU-MIMO gain.

# 6. Wi-Fi Application Scenario

In the last years, Wi-Fi has become the most widespread wireless local area network technology worldwide. It is nearly supported by all user communication devices and can be found in many deployment scenarios, both private and enterprise. Since the first release of the Wi-Fi IEEE 802.11 standard in 1997, the technology has rapidly improved by increasing the data rate to multiple Gbit/s and supporting many more users per deployment [5], [72]. However, the dramatic increase in Wi-Fi network and user density created problems of its own. The dense deployment results in a shortage of available channels and inter-user contention and interference [7]. As written in Chapter 1, one key technology leap to overcome this problem was the introduction of the MU-MIMO operation in 2013 with IEEE 802.11ac [5], [6]. Since 2013, the adoption of the MU-MIMO technology was quite slow and current commercially available devices do not achieve the full capabilities as defined in the standard [8]. Given current hardware constraints, the literature [73]–[75] and also vendors of commercial Wi-Fi solutions like Qualcomm [8] concluded that, in typical channel conditions, the gain of MU-MIMO is relatively limited. In this chapter, we address the problem of the low MU-MIMO performance in Wi-Fi networks. We propose the HDA architecture as a solution to increase the number of antenna elements. The lower complexity of the hybrid approach is an important factor for commercial Wi-Fi devices. The reduced cost should make the solution feasible for Wi-Fi. In addition, the goal is to find a solution based on COTS Wi-Fi hardware, that does not require a protocol change as this would require a new standard release or some proprietary solution.

We demonstrated the proposed solution in the demo session of ACM MobiCom 2020 [38]. The following chapter is based on an original journal manuscript [39]. The journal manuscript will be submitted to the IEEE/ACM Transactions on Networking.

## 6.1. Problem Statement and Related Work

In the background chapter in Section 2.1, we explained the interconnection between the capacity of a MU-MIMO system, the number of antenna elements/number of spatial streams, and the power penalty of the precoding. If we want to increase the capacity, we can increase the number of antenna array elements at the AP to increase the number of spatial streams.

The Wi-Fi standard sets a maximum number of spatial streams and limits this possibility. Reducing the power penalty of the precoding is the second option to increase the capacity.

The power penalty of the MU-MIMO precoding depends on the quality of the channel. For ZF, it is inverse proportional to the condition number of $\mathbf{H}$ as defined in (2.4). The exact relation between the power penalty and the channel might be different for other algorithms. Nonetheless, the power penalty will still depend on the quality of the channel. The specific algorithm and its implementation used in a Wi-Fi chip are often unknown. Independently of the exact precoding algorithm, the Wi-Fi system has to select a set of users for the MU-MIMO precoding. This selection is also called MU-MIMO user grouping or scheduling. The set of users does specify the effective channel $\mathbf{H'}$ and the $\alpha'_k$ coefficients for the selected users. The algorithm can select all users, in which case $\mathbf{H'} = \mathbf{H}$, or a subset of the users. In case of an ill conditioned channel, it might be optimal for the Wi-Fi AP to serve less users than the maximum possible. The effective channel $\mathbf{H'}$ with the dimension $N' \times M$ for the smaller set of users ($N' < N$) will be better conditioned (i.e., the $\alpha'_k$ coefficients will be larger) than the full channel. As a result, the sum data rate will be higher. Similar to the precoding algorithm, the user selection algorithm is not known for commercial Wi-Fi systems.

To boost the MU-MIMO performance of a system based on COTS Wi-Fi, the Wi-Fi precoding and user selection algorithms should be enabled to serve the maximum number of users. This can be done by altering the channel as it is seen by the Wi-Fi system. The power penalty for all users should be as small as possible to support high modulation schemes when all users are selected.

*Related Work:* The problem of ill-conditioned wireless channels for Wi-Fi networks has seen some interest in the last couple of years. Different approaches have been investigated to increase the single-user MIMO or MU-MIMO performance. Beside the research on Wi-Fi systems, a lot of literature exists on general MU-MIMO systems. These publications are not based on COTS Wi-Fi hardware and the current Wi-Fi protocol. Hence, we can not compare to them. In the following, we will only list Wi-Fi-related literature.

Most of this literature focuses on relay based networks or cooperating APs. In [76], the authors describe a full-duplex relay system. It is based on single antenna relays and does only consider single-user MIMO. Another system [77], called megaMIMO 2.0, uses a collaborative network of APs to create additional propagation paths to STAs. It requires reference channels between the APs to exchange the user data and reference information. The necessary synchronization between the APs can not be implemented without Wi-Fi protocol changes. The more recent works RFocus [74], LAIA [78] and ScatterMIMO [75] all propose to alter the wireless channel with reflecting arrays. RFocus [74] uses a passive switch based antenna array to enhance the channel to a single user at a time and does not

increase the MU-MIMO performance. The LAIA [78] system employs a passive antenna array as relay to improve the channel conditions for MU-MIMO. It requires the channel knowledge of the STAs at the relay. Hence, it introduces additional overhead for feedback and needs Wi-Fi protocol changes. The latest work ScatterMIMO [75] introduces smart surfaces which are reflecting antenna arrays as relays. The smart surface adds artificial paths to the wireless channel of a single STA. It requires the CSI of the AP at the smart surface and works without modifications of the Wi-Fi protocol. The authors demonstrate ScatterMIMO with COTS Wi-Fi hardware in a single-user MIMO operation. They also propose to use multiple smart surfaces for MU-MIMO operation.

Another part of the literature does not use relay networks but focuses on the AP hardware. The Phaser [79] system uses multiple combined COTS APs to increase the number of antenna elements. The authors introduce a calibration scheme so that the APs can be used as a single system. As described in section Section 6.1, this increases the maximum MU-MIMO gain but not the orthogonality between the user channel vectors. Following, the system might still not reach this maximum due to the channel condition. The more recent SWAN [73] approach connects multiple antenna elements with switches to the RF chains of a COTS AP. This might improve both the maximum MU-MIMO gain and the channel condition. In [73] the authors only consider single-user MIMO. In general, an extension to MU-MIMO might be possible but was not investigated or demonstrated in [73].

## 6.2. Hybrid Beamforming for Wi-Fi

As pointed out in Chapter 2, the fully-digital massive MIMO approach, using much more antenna elements than users, solves the problem of ill-conditioned channels. However, the cost makes it infeasible for the use in Wi-Fi networks. As we have seen in Chapter 3, the HDA architecture is a solution to cost and complexity issues. In this chapter, we propose to use HDA beamforming to achieve the full MU-MIMO capabilities of the IEEE 802.11 standard. The analog hardware is forming an effective channel for the digital Wi-Fi system, that is more diagonally dominant and, thus, has a lower condition number increasing the average MU-MIMO gain with respect to its maximum. Compared to the fully-digital massive MIMO approach, the hardware complexity of an HDA system is lower and makes it feasible for COTS Wi-Fi. We developed a demonstrator to evaluate the concept and show the performance gain compared to standard Wi-Fi. The user stations are COTS Wi-Fi equipment. We measure the real user throughput on the application layer in the downlink including the protocol overhead. As a benchmark, we compare the performance of our HDA system with an unmodified COTS four element antenna system in different scenarios.

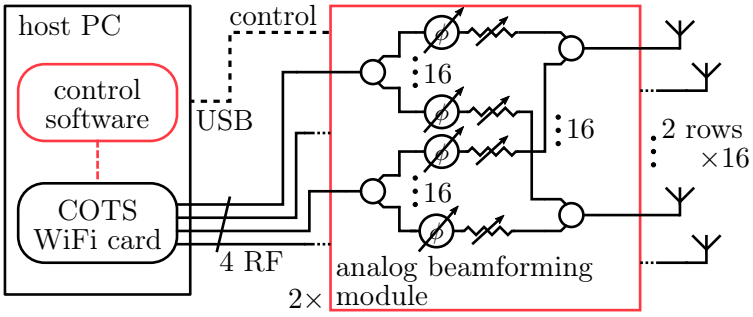



Figure 6.1.: Architecture of the HDA MU-MIMO Wi-Fi system with the HDA extension marked in red.

### 6.2.1. Hybrid Beamforming Concept

We propose to use HDA beamforming to achieve the full MU-MIMO capabilities of the IEEE 802.11 standard. The analog beamforming is steering beams towards the STAs which is equivalent to forming an effective more diagonally dominant channel for the digital system. The digital system is based on a COTS Wi-Fi system. The analog network reduces the channel dimension seen by the digital Wi-Fi processing from $M$ to $M_{\mathrm{RF}}$. As stated in Section 6.1, we can increase the MU-MIMO gain by improving the channel quality (i.e., decreasing the condition number of the channel matrix). This holds for both the full communication channel $\mathbf{H}$ and the effective channel $\tilde{\mathbf{H}}$. Following from $M_{\mathrm{RF}} < M$, the maximum spatial gain of an HDA system is $M_{\mathrm{RF}}$, requiring $K \leq M_{\mathrm{RF}}$.[1] The analog beamforming forms a diagonally dominant $\tilde{\mathbf{H}}$ for the digital system. Hence, the condition number of $\tilde{\mathbf{H}}$ is smaller, and as a result the MU-MIMO performance is better, compared to a channel given by a system with $M_{\mathrm{RF}}$ antenna elements directly sounding the physical channel. Current commercial Wi-Fi systems are such systems, directly probing the channel with each RF chain. As mentioned in the introduction, those systems do not achieve the maximum spatial gain $K \sim M_{\mathrm{RF}}$ (e.g., as described in [8]). Using our demonstrator, in the following sections, we show that an HDA system increases the MU-MIMO performance compared to a COTS Wi-Fi system. We achieve $K = M_{\mathrm{RF}}$ with high probability.

We present an HDA architecture suited for the Wi-Fi protocol. In order for the HDA beamforming to work with the Wi-Fi protocol and COTS hardware, the concept described in Chapter 3 needs to be adapted. We adapted the HDA-specific signal processing, introduced in Section 3.2, for the Wi-Fi sub-6 GHz use case. The hardware uses an custom analog part and a Wi-Fi part. The analog part consists of an antenna array and an analog beamforming module. The analog beamforming uses the analog module described in Chapter 4. It

[1] Of course the system is also limited by the maximum number of user streams $K \leq N$.

provides a beamforming gain and forms the improved effective channel. The digital part is a COTS Wi-Fi network interface card (NIC) embedded in a host PC. The RF chain ports of the analog network are connected to the RF ports of the Wi-Fi card. COTS Wi-Fi devices are closed systems without many controllable parameters. Hence, we cannot choose and even do not know the algorithms used for the digital MU-MIMO precoding and MU-MIMO STA grouping. The digital part is like a black box for the system. The host PC also runs the control and signal processing software necessary for the HDA approach. Figure 6.1 shows an overview of the system. The hybrid concept is fully transparent for the digital processing, i.e., no changes to the Wi-Fi protocol are needed. The advantage of this system is, that it is fully compliant to the Wi-Fi protocol since the digital part is controlled by the COTS Wi-Fi hardware. The disadvantage is, that the analog beamforming and the digital precoding can not be jointly optimized. Nonetheless, the beamforming and precoding algorithms described in Section 3.2.2 are not jointly optimized and show a very good performance. Hence, for our HDA Wi-Fi system, we use the proposed HDA precoding. The analog network steers the signal of a RF chain towards the strongest path of a single STA with a beam pattern calculated by the standard phased array (Fourier-matrix based) algorithm [54]. The analog beamforming is transparent to the Wi-Fi hardware and creates an effective channel with a lower condition number. Following, even an unknown digital precoding algorithm should in general have a larger MU-MIMO gain due to the better effective channel.

The analog beamforming and the MU-MIMO user grouping, in contrast to the precoding, should be jointly optimized. The analog beamforming does change the effective channel seen by the MU-MIMO user grouping of the Wi-Fi chip. It influences the Wi-Fi user selection and, to a certain extend, enforces a user set. However, the user grouping is not only dependent on the channel but also on other parameters like fairness and traffic demand. Hence, the HDA control system has to select appropriate users. The HDA system does need to steer beams towards STA which have traffic and need to be included in the MU-MIMO user set. The HDA system either would have its own user grouping based on the traffic information and the directions of the users or would cooperate with the Wi-Fi user grouping system. For the former case, the HDA system would acquire the traffic information from the Wi-Fi system, e.g., by tapping into the Wi-Fi driver. The latter case would require a Wi-Fi STA grouping which is aware of the HDA system. An optimized commercial HDA Wi-Fi system would most likely have a co-design of the HDA signal processing and the Wi-Fi chip/driver. The MU-MIMO grouping could, for example, not only be based on the user traffic but also on the directions of the users to increase the spatial separation. We did not investigate such algorithms. The literature already proposed some algorithms, also called spatial scheduling algorithms, but mainly for mobile communication [19]. The

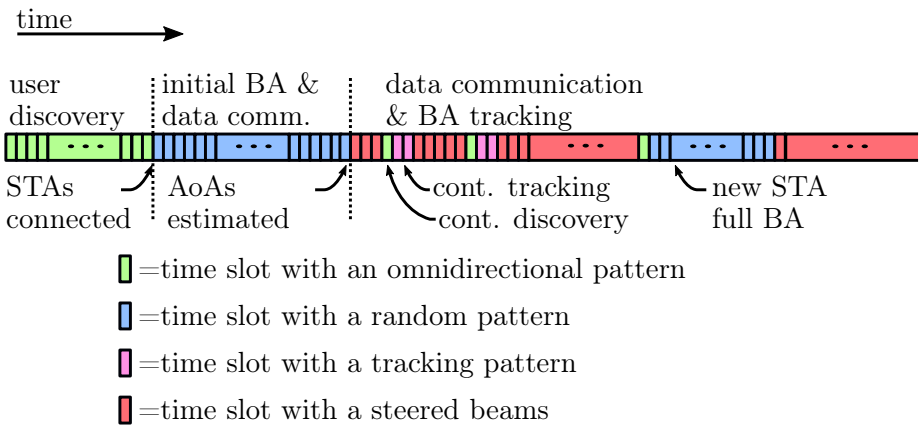



Figure 6.2.: Time sequence of the HDA Wi-Fi control protocol.

scheduling and the MU-MIMO user grouping problem are interesting topics for future research especially with respect to Wi-Fi. Our testbed always operates with the same number of user streams and AP RF ports and, hence, does not require any change to the Wi-Fi STA grouping. An extension of our testbed with a separate user grouping for the HDA beamforming would be possible by acquiring the traffic information from the network stack of the AP, e.g., the driver of the COTS Wi-Fi NIC.

The analog hardware is reciprocal and steers beams for both the downlink and the uplink. The described downlink concept holds for uplink MU-MIMO as well. Unfortunately, most currently available COTS Wi-Fi hardware does not yet support the newest IEEE 802.11ax standard and MU-MIMO in the uplink. This includes the Wi-Fi NIC that we use for the AP in our demonstrator. The analog beamforming during the uplink does increase the SNR for a single stream user device but might reduce the single-user MIMO performance.

In general, HDA systems can be used for many different WLAN application scenarios. Nonetheless, they are most practical for Wi-Fi deployments in large rooms and with many users, i.e., larger cells are more beneficial. The increase of the number of antenna elements of course also increases the physical size of the antenna. Also, as discussed, the MU-MIMO performance of the network is limited by both the number of spatial streams at the AP $K$ and at all users $N$. For a network to reach the limit, it needs at least $N \geq K$ user streams.

### 6.2.2. Control Protocol Design and Signal Processing

To realize the HDA concept, besides the hardware also some form of protocol controlling the analog beamforming is necessary. Our concept works on top of the Wi-Fi protocol and operates fully transparent to the Wi-Fi stack. The Wi-Fi protocol runs on the NIC and the HDA protocol as software on the host PC. Of course, in a more integrated system

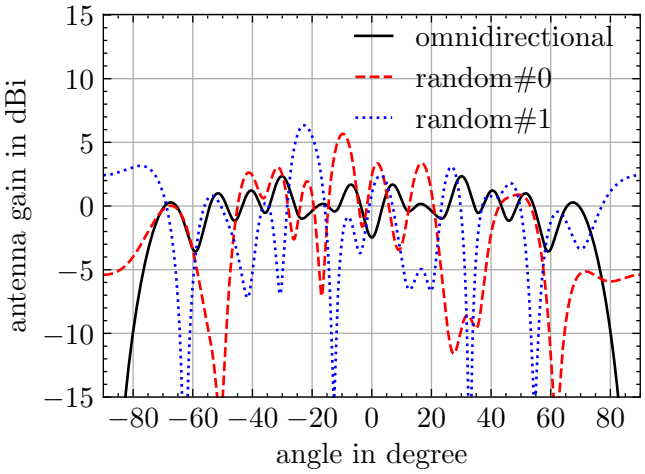



Figure 6.3.: Example antenna array gains for a single RF chain for the discovery phase (omnidirectional pattern) and the BA phase (two random phase patterns).

the HDA control could be included in the Wi-Fi stack. The time flow of the protocol is depicted in Figure 6.2. The time is divided into slots with a certain length. This length is a configuration parameter and for our demonstrator in the range of 10 ms to 100 ms. The analog beamforming configuration can be changed between the slots and is constant within each slot.

The protocol starts without connected STAs. It sets up the analog beamforming with a quasi-omnidirectional beamforming pattern for all RF chains. This pattern is constant for all discovery time slots. The array gain of the pattern is shown in Figure 6.3. This pattern has roughly an antenna array gain of 0 dB in all directions. This guarantees that users can be discovered in all directions. After a STA is connected to the AP, the data connection is established and works similar as with a normal AP. This initial phase is called user discovery in Figure 6.2. After at least one STA established a connection, the protocol starts the next phase.

The next phase is the initial beam alignment (BA). The AP estimates the directions of the strongest paths for all connected STAs using the proposed BA algorithm. The algorithm is described in detail in Section 3.2.1. We adapted the algorithm to the Wi-Fi use case. Contrary to the standard version of the algorithm from Section 3.2.1, the adapted algorithm uses measurements from the uplink traffic received at the AP. The estimation in the uplink is even required because the BA algorithm running on the AP does not have access to measurements on the COTS Wi-Fi STAs. The STAs are unaware of the HDA system. One limitation due to the uplink measurements is, that STAs should not employ beamforming in the uplink. An additional beamforming direction on the STA side would

require a joint estimation of both the AP and STA direction. Since the beamforming hardware is reciprocal, the estimation based on the AoA in the uplink can be used as AoD information for the precoding in the downlink.

The adapted BA uses different beamforming patterns than described in Section 3.2.1. The AP sets up the analog beamforming with a random phase and an amplitude of one per element. The phases of the elements are randomly changed for each time slot. The antenna gain of two example beam patterns is shown in Figure 6.3. These patterns have on average over multiple time slots an antenna array gain of 0 dB in all directions so that connected STAs do not suffer from long connection losses. This would not be the case with the standard BA patterns as described in Section 3.2.1. In addition, due to the lower path loss in the sub-6 GHz Wi-Fi frequency range compared to the mm-wave range, we can do without the additional array gain of the multibeam patterns. The connection quality is similar to the initial discovery phase. The BA routine measures the received power per STA for each pattern. It uses the received signal strength indicator (RSSI) value reported by the Wi-Fi driver.[2] Each RF chain can be used to measure a different pattern. Following, four measurements are taken per training slot. The normal uplink traffic is sufficient and no special training packets are required. Once a large enough number of different patterns was measured, the algorithm computes the AoAs of the strongest paths for all connected STAs as formulated in Section 3.2.1. The number of required trainings slots depends on the channel condition. It is a compromise between the training overhead and the detection probability. The number of required time slots for the BA is independent of the number of available users. The directions of all users are estimated at the same time by the AP. Once the directions are estimated, the protocol stores the AoAs of the STAs into a database and continues with the following phase.

The final phase is the MU-MIMO data communication phase. The protocol sets up the analog network to steer beams towards the STAs to increase the MU-MIMO gain. As described in Section 3.2.2, the signal of each RF chain is steered with a single beam towards an AoA of a single STA. The antenna gain of such beam patterns is shown in Chapter 3 in Figure 3.4. The beam patterns can be changed or held constant between the time slots. Depending on the number of connected STAs and their total number of possible spatial streams $N_{\text{avail}}$, the protocol either steers beams constantly to all users (in case of $N_{\text{avail}} \leq K$)[3] or to the user set selected by the MU-MIMO user grouping mentioned in the last section (in case of $N_{\text{avail}} > K$)[4].

[2] which is the received power in dB for Qualcomm and Intel Wi-Fi NICs

[3] The number of user streams is less or equal to the number of AP spatial streams.

[4] The number of user streams is higher than the number of AP spatial streams; not all users can be served constantly.

During the communication phase, the protocol needs to track the directions of the users to support moving users. We have implemented a simple tracking algorithm . The tracking periodically schedules time slots with different patterns. The frequency of the tracking slots is a configuration parameter. It depends on different system parameters like the width of the beams[5] and the speed of the moving users. During each tracking slot, the AP tests adjacent beam directions around the current beam direction of each user. The directions are chosen from a grid with a distance of 3° between the points. The angular distance of 3° is selected in such a way, that the beams are overlapping. Due to the overlapping beams, the probability of a connection loss or a drop in the data rate for the users is greatly reduced. As during the BA phase, the AP measures the received power for the tracking slots. The direction of the slot with the highest power is declared the new current beam direction. We have not optimized the tracking behavior and algorithm.

Besides the tracking, the HDA protocol also needs to schedule time slots for the discovery of new users and management frames, e.g., beacons. During these time slots the protocol sets up the same quasi-omnidirectional pattern as in the discovery phase. It assigns this pattern only to a single RF chain per slot. This is sufficient and necessary so that the AP can discover users in all directions. The other RF chains are still steered towards scheduled users. Hence, even during the discovery slots the MU-MIMO operation can continue. In Figure 6.2, we have drawn the periodically tracking and discovery slots side by side. This is a simplification, both the period and the exact timing can be different. In case a new user established a connection during a discovery slot, the protocol would run a full BA cycle. This BA estimates the AoAs of the strongest paths for the new and all previously connected STAs.

As written before, this protocol and the beamforming are completely transparent for the Wi-Fi stack. The beamforming changes the effective channel seen by the Wi-Fi system. Nonetheless, during all times all users can communicate with the AP, under the condition of the effective channel. The Wi-Fi protocol can choose the modulation and coding scheme (MCS) with both single-user MIMO and MU-MIMO operation. The goal of the HDA protocol is to improve the effective channel for the users with a traffic demand in a way, that the Wi-Fi system selects the MU-MIMO coding scheme.

## 6.3. HDA-Wi-Fi Demonstrator

We build a demonstrator to investigate the proposed HDA MU-MIMO concept for Wi-Fi. Our implementation of the system consists of hardware and software components. Figure 6.1

[5] which determines how far a user can move without a decrease of the received power

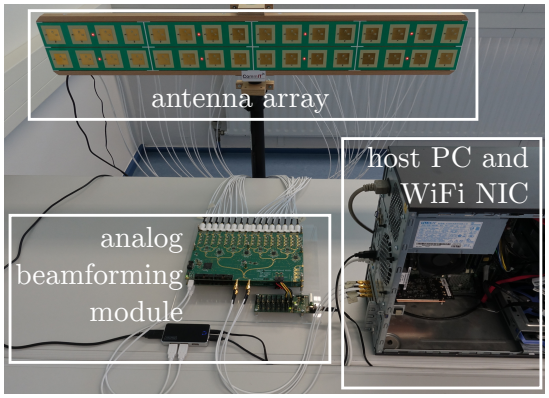



Figure 6.4.: A picture of the hardware of the HDA-Wi-Fi demonstrator.

shows an overview of the system and Figure 6.4 a photo of the hardware. The main components are the antenna array, the analog beamforming module, the COTS Wi-Fi NIC, and the control and signal processing software running on the host PC. The demonstrator is a fully functional 802.11ac AP. It can be used with COTS user stations. We have presented the demonstrator during ACM MobiCom 2020 [38]. In this section, we describe the implementation details with respect to the described concept.

### 6.3.1. Hardware Design

The main hardware components are the self-developed antenna array and analog beamforming module and the COTS Wi-Fi NIC. The antenna array and the analog module are introduced in Chapter 4. In this chapter, we use an array configuration with 2 rows and 16 columns and a vertical polarization. For the analog network, we use two analog modules in parallel to connect the four RF ports of the Wi-Fi hardware to the two rows of the antenna (see Figure 6.1). This HDA structure is a mix between the FC and the OSPS architecture. Following, the rows are independent ULAs and subarrays of the whole antenna. One could call this architecture two-streams-per-subarray. Because of this structure, the beamforming is restricted to the azimuth dimension and elevation beamforming is not possible. We chose this structure to reduce the complexity and the insertion loss due to the power divider network.

For the digital precoding and Wi-Fi processing, we use the COTS 802.11ac NIC QNAP® QWA-AC2600 with 4 RF ports. The card can simultaneously operate in the 5 GHz and the 2.4 GHz band. We only use it in the 2.4 GHz band. This card is based on a Qualcomm® QCA9984 chip, which supports IEEE 802.11ac with MU-MIMO and 4 antenna elements. Unfortunately, we found out that in our configuration even under perfect channel conditions

(i.e., stations connected with RF cables) this chip only supports three concurrent single stream users in MU-MIMO mode. This limitation might be a design choice by Qualcomm [8]. However, the chip supports four data streams for non-single stream users (e.g, two users with two streams per user). We have included measurements in Section 6.4.2 showing these limitations. The card can use a modulation up to 256-QAM. The maximum MCS is 8 in the 2.4 GHz band. The card is plugged into a standard Intel processor-based PC.

### 6.3.2. Software and Signal Processing Implementation

The software runs on the host PC with a GNU/Linux operating system and the kernel in version 5.4. The COTS Wi-Fi card uses the default unmodified ath10k driver and firmware. Unfortunately, the driver does not have a control interface for the transmission mode (MU-MIMO or single user) or any MU-MIMO parameter (e.g., user selection and grouping). We can only influence the transmission mode by altering the effective channel so that the Wi-Fi chip selects the MU-MIMO operation. However, a fixed MCS index can be set. This helps in testing and measuring the MU-MIMO performance because it effectively sets a signal-to-interference-plus-noise ratio (SINR) requirement. We run hostapd[6] to enable access point functionality towards client stations. The user data path is fully decoupled from the control software of the HDA system, except the MCS setting used for testing.

The HDA control software is written in Python. The control interface of the analog beamforming module is connected via USB to the host PC. The reconfiguration of a beamforming pattern requires around 5 ms.[7] We implemented all protocol parts (except the MU-MIMO user grouping) and the BA algorithm as described in Section 6.2. The time slot duration of our implementation in contrast to the concept description is not fixed. The discovery phase starts after enabling the demonstrator and the user STAs. The connections are established after a couple of seconds mainly determined by the software runtime. The control software initiates uplink traffic from the users for the BA phase by running the `ping` command. The interval of the `ping` command is set to 10 ms. This results in an average training time slot duration of 20 ms. This includes the beamforming setup time and the time until the protocol acquired RSSI measurements for all STAs. Through experimentation, we determined the number of required time slots for the estimation algorithm to be 64. The results of the BA performance evaluation are given in Section 6.4.2. The overall time required by the BA to estimate the AoAs of four users is about 1.4 s which is already sufficient for the envisioned indoor scenario with no or low mobility.

[6] `https://w1.fi/hostapd/`
[7] Which is mainly determined by the USB interface and could be shortened.

Because of the missing direction aware user scheduler and an interface to the Wi-Fi chips STA grouping, we use a fixed set of users for the data communication phase tests. Following, we do not change the beamforming patterns during the data communication phase except for the tracking time slots. The tracking period is 1 s. During each tracking, the AP measures in three time slots the received power of all users for their current direction and the two adjacent directions on the 3° grid. The tracking also initiates uplink packets using the `ping` command. In Section 6.4.2, we show measurements on the influence of the tracking on the data rate performance.

## 6.4. Performance Evaluation

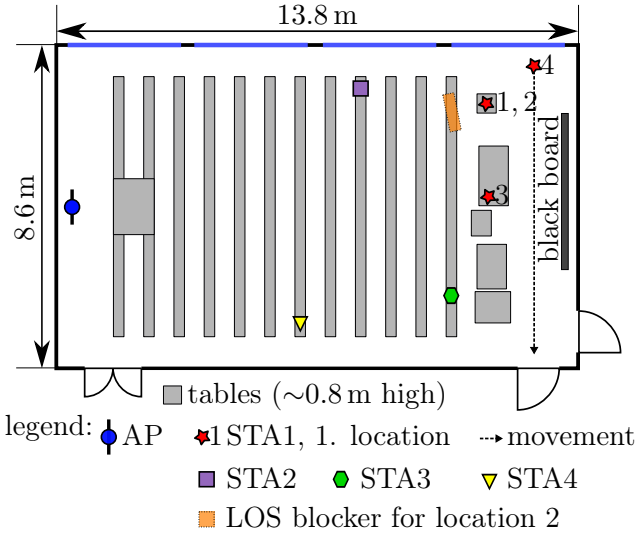
We evaluate our proposed HDA concept with a full system evaluation of our demonstrator as well as with microbenchmarks of certain key aspects of the concept. The first step to understanding the concept and its advantages is to measure the effective channel seen by the Wi-Fi system (i.e., the digital precoding). As described in Section 6.1, the condition number of the effective channel does relate to the expectable MU-MIMO performance. To calculate the condition number, we measure the CSI (i.e., the complex coefficients of the wireless channel for each sub-carrier) between every RF chain. This provides insights into the influence of the HDA concept on the effective channel. Since COTS Wi-Fi systems do not support the measurement of the MU-MIMO effective channel, the CSI measurements are not based on the COTS Wi-Fi system but on a SDR based system. The second part of the evaluation is to evaluate the actual system implementation. We measure the BA performance, the tracking system and the overall throughput performance. These measurements are acquired with the described demonstrator based on the COTS Wi-Fi system. First in this section, we explain the general evaluation methodology and the two systems, one for the CSI measurements and one for the throughput evaluation. Second, we present the microbenchmarks and the end-to-end results from the measurements

### 6.4.1. Evaluation Methodology and Setup

#### General Evaluation Setup

All measurements are performed in Wi-Fi channel 1. The center frequency is 2.412 GHz, and the bandwidth is 20 MHz. The results scale with the bandwidth. In general, the HDA concept can be used with a larger bandwidth. The channel was unoccupied by other Wi-Fi networks during the measurements. The measurements were obtained in a large lecture hall. Figure 6.5 (a) is a picture of the location and the measurement setup. The map in Figure 6.5 (b) shows the dimensions of the room and the locations of the AP and the


(a)




(b)

Figure 6.5.: The measurement environment: (a) A picture of the room. (b) A map showing the dimensions and the user locations.

STAs. The AP antenna array is positioned in a height of 2.2 m. The STAs stand on tables in a height of 0.8 m. The room is in the back 4.1 m high and has a slope from the back, where the AP is positioned, to the front with a height difference of ∼0.8 m. The locations of the AP and the three STAs 2, 3, and 4 are fixed. The possible locations of STA 1 are marked in the map and named accordingly in the results section. Some measurements were acquired with STA 1 moving along the given path marked as location 4 in Figure 6.5 (b). The location of STA 1 for the location settings 1 and 2 is the same but in case of 2 the LOS is blocked. We block the LOS with a large piece of radiation-absorbent material directly in front of STA 1 as shown in Figure 6.5 (b).

We use a four-element antenna configuration as a baseline for both the data rate and the CSI evaluation. The four-element antenna is the standard antenna bundled with the QNAP® QWA-AC2600 card. We use four RF switches to connect both the HDA hardware and the four-element antenna to the four RF chains of the Wi-Fi card. Hence, we can compare both antennas in the exact same measurement scenarios; for example, at the same user locations.

Due to the high insertion loss of the analog module, we decided to add four 12 dB attenuators between the four-element antenna and the RF chains for compensation. Thus, the HDA system and the four-element antenna have roughly the same total output power.

### CSI Measurements

We want to evaluate not only the final system performance but also the HDA concept by investigating the effective channel seen by the Wi-Fi system. To do this, we measure the CSI at the user side for the full composite effective channel, in our case $\tilde{\mathbf{H}} \in \mathbb{C}^{4\times4}$. The COTS Wi-Fi hardware of our demonstrator does not provide this data.[8] Hence, instead of the COTS Wi-Fi AP and STAs we use two of our self-developed SDRs named ExsV.[9] One ExsV offers four RF chains with a frequency range of 0.7 GHz to 3.0 GHz and a maximum bandwidth of 30 MHz. Of course, the RF parameters, like the frequency stability, of a SDR and a COTS NIC are different. Nonetheless, the relative change in the measurements allows us to analyze the MU-MIMO performance.

We connect one SDR to the HDA hardware and the four-element antenna at the AP side. The other ExsV is connected to the omnidirectional antennas of the four STAs using long RF cables of the same length. All antennas are at the exact same locations as during the throughput measurements. The SDRs are not synchronized both in time and frequency. The SDR at the AP transmits continuously (with some gap) Wi-Fi null data packets (NDPs)

[8] different NICs provide measurements of the CSI per data stream but none for the full channel

[9] `https://www.commit.tu-berlin.de/menue/forschung/testsysteme_und_prototypen/exsv_versatile_4_channel_sdr_platform/parameter/en/`

as defined in IEEE 802.11ac. NDPs are used in the MU-MIMO mode of Wi-Fi to estimate the CSI. The ExsV at the STA side acquires enough samples so that at least one full packet is received. The received packets are recorded on a host PC. The signal processing to estimate the CSI is done afterwards in non-real time. We use standard algorithms to receive the packets and to estimate the CSI. The packet detector correlates with the legacy long training field (L-LTF). The frequency offset is also corrected on the basis of the L-LTF. Afterwards, we estimate the full CSI using the very high throughput long training field (VHT-LTF). From the CSI, we derive the condition number of the effective channel. We use MathWorks® MATLAB WLAN toolbox for both the NDP generation and the CSI estimation.

### COTS Throughput and BA Evaluation

The throughput and the BA performance evaluations are based on the demonstrator as described in Section 6.3. The user STAs are small computers running a GNU/Linux operating system. The Wi-Fi NICs are Intel® Wireless-AC 9260 cards, which support IEEE 802.11ac with up to two streams. Since we are mostly interested in the MU-MIMO performance, we connect only one omnidirectional antenna to each card and disable the second stream.[10] Each measurement starts with the initial user discovery. The STAs are connected to the Wi-Fi network of the AP as in any COTS network.

The BA performance evaluation is based on the demonstrator protocol implementation but records the RSSI measurements. Also, during the BA evaluation the number of random pattern training slots is larger (i.e., 512 slots). The performance characterization is calculated afterwards using the recorded data. We run the BA algorithm for multiple subsets of the 512 measurement values with different sizes. We compare the estimated AoA with the known location. We calculate the detection probability $P_D$ (i.e., the probability of finding the correct AoA $\pm 3°$).

The downlink throughput evaluation of the Wi-Fi basic service set (BSS) uses the full demonstrator setup. After the user discovery, the standard BA with 64 trainings slots estimates the AoA. Afterwards in the data communication phase, the throughput is measured using the user data path in the downlink. The downlink throughput of the network connection is evaluated using the iperf3[11] program. Each STA runs an iperf3 server. The AP starts one iperf3 client process per STA. The used protocol is UDP and the packet size 1448 byte. The measurement duration of a single data point is 1 s which is the shortest possible duration with iperf3. iperf3 runs without a bandwidth limitation and

[10] by changing the capability settings of the NIC
[11] `https://software.es.net/iperf/`

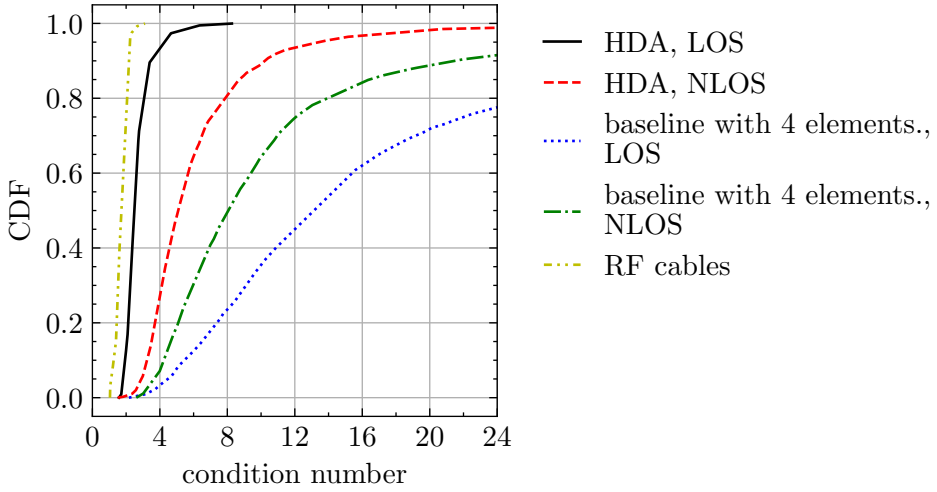



Figure 6.6.: CDF of the condition number of the CSI of the HDA system and the baseline system. Shown for the LOS (location setting 1) and the NLOS (location setting 2) case.

constantly fills the transmission buffers of the AP NIC. Hence, the maximum throughput is determined by the Wi-Fi card. The iperf3 server on the user side measures the throughput taking lost packets into account. It reports the throughput back to the AP where we record it.

During the measurements, we can change the set of active iperf3 clients on the AP. This changes the set of users the Wi-Fi card needs to serve. We can also fix the used MCS index or allow all possible indexes. Both settings help us to better evaluate and compare the HDA and the four-element antenna system.

### 6.4.2. Results

In the following, we present the microbenchmarks and the end-to-end results from the measurements. We begin with microbenchmarks of some key aspects of the proposed HDA Wi-Fi system. These include the condition number of the effective channel, the BA performance, and a validation of the measurement setup using RF cables. Afterwards, we show the results of a full end-to-end system evaluation of our demonstrator with different location and MCS setting scenarios.

#### Effective MU-MIMO Channel

To understand the effect of the HDA concept, we investigate the effective channel as it is seen by the digital signal processing. We use the condition number as the performance

measure of the channel. We compute it for each subcarrier. The smaller the number, the better the channel is suited for MU-MIMO. The measurement setup is described in Section 6.4.1. We compare the HDA system with the baseline four-element antenna. The cumulative distribution function (CDF) of the condition number is shown in Figure 6.6. The CDF includes the measurements of all subcarriers. We plot the curves for a location with a LOS condition (location setting 1) and with a NLOS condition (location setting 2). As a baseline, we added the measured condition number when the two SDRs are connected directly with 50 dB attenuators and RF cables. This baseline has a mean of 1.68. In an ideal setup, the condition number would be 1. This difference can be explained by inaccuracies in the estimation, due to detection and frequency offset errors, and nonideal hardware properties (e.g., coupling in the devices). The SNR of all measured packets in all cases was between 26 dB and 28 dB.

The HDA system does achieve a much better condition number than the baseline system. The mean in the LOS case of the HDA system is 2.57. It is by a factor of 7.3 better than the value with the baseline antenna, which is 18.76. In the NLOS case, the condition number of the HDA system becomes much higher and the difference to the four-element antenna smaller. The mean of the HDA system is 6.34 and of the four-element antenna 12.15. It is interesting to see, that the condition number of the four-element antenna is smaller in the NLOS case compared to the LOS case. The HDA system is much more dependent on the LOS, which is in any way clear because the system does point the beams towards the strongest paths. Nonetheless, the HDA system does always form a better effective channel for the digital signal processing.

#### BA Performance

The previous measurements proved that the HDA concept can increase the MU-MIMO performance of a Wi-Fi system. But to achieve this, the system does need to know the correct AoAs of the STAs. We evaluate the performance of the proposed BA scheme using the demonstrator and the setup as described in Section 6.4.1. Figure 6.7 shows the detection probability $P_D$ over the number of training slots. We calculate the probability of the correct AoA estimation for multiple measurements at the fixed location settings 1, 2, and 3. $P_D$ converges to $\sim$1 between 20 and 30 training slots. Following, with our decision to use 64 training slots we should always estimate the correct AoA and have some security margin.

#### RF Cable Benchmark

To get an understanding of the best possible performance with the selected Wi-Fi hardware, we measured the end-to-end throughput using RF cables. We connected the four RF chains

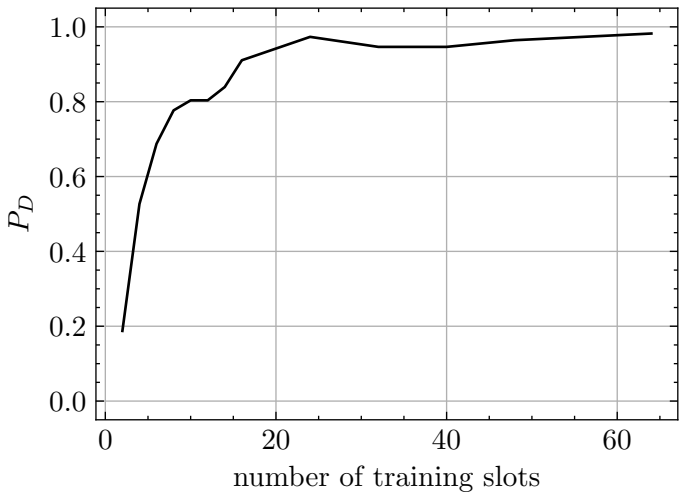



Figure 6.7.: The initial BA detection probability versus the number of training slots. It combines multiple measurements at the location settings 1, 2, and 3.

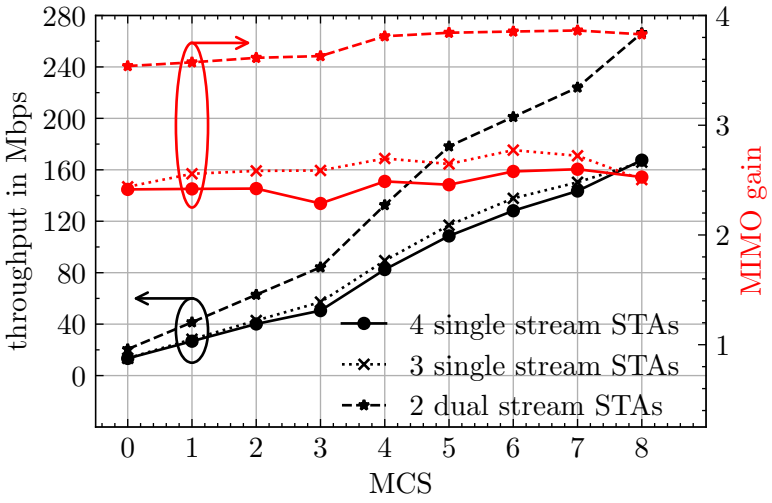
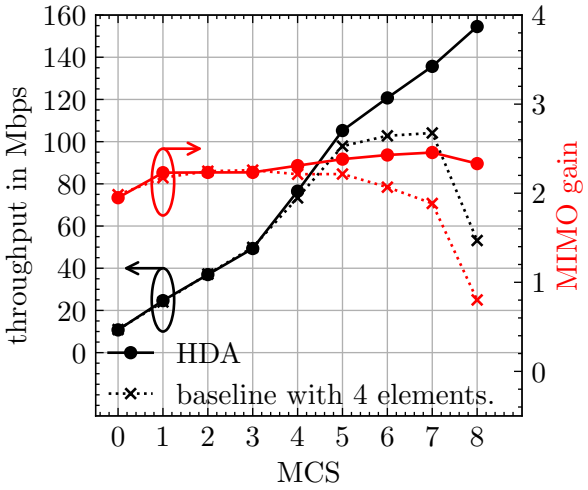



Figure 6.8.: Mean of the throughput and the MU-MIMO gain for each MCS measured with the STAs connected by cables to the AP.

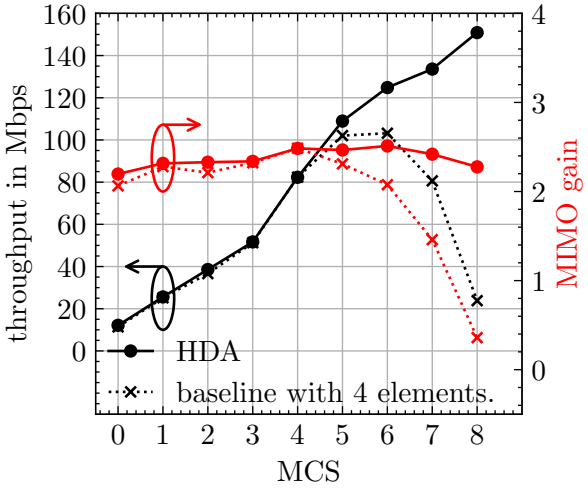
of the AP directly with 50 dB attenuators and RF cables to the four STAs. The UDP throughput for all possible MCS settings can be seen in Figure 6.8. We also plotted the achieved MIMO gain on the second axis. We calculated the MIMO gain by dividing the measured throughput by the optimal achievable throughput for each MCS. As we previously mentioned in Section 6.3.1, the Qualcomm® QCA9984 chip does only support three spatial streams for MU-MIMO. The single stream curves are constantly below 3 (between 2.4 and 2.8). This deviation from 3 might be due to the channel sounding overhead of IEEE 802.11ac where compressed beam (CB) frames need to be send in uplink. We measured two different setups for single stream STAs. In the first setup, we enabled only three STAs. The second setup uses all four STAs. The four STAs have a slightly lower performance than the three STAs. The CSI measurement overhead for four STAs is larger than for three, so this was expectable. In addition to the single stream STAs, we also measured the throughput with two dual stream STAs. In this case, the system achieves a MIMO gain of up to 4. Since we propose the HDA concept to increase the MU-MIMO performance, we will continue with the single stream user scenario. For this, the maximum measured MU-MIMO gain was 2.8 and the maximum throughput 175 Mbps.

#### End-to-End Throughput performance

After we have validated the concept by investigating the effective channel, measuring the BA performance and learning the limitations of the Wi-Fi hardware, we are in the following presenting the over-the-air end-to-end downlink throughput results. The results are measured with the system as described in Section 6.4.1. For the fixed location settings, we have disabled the tracking system to measure the optimal performance. The tracking system is not yet fully optimized. In Figure 6.9, we show the throughput for all possible MCS settings. As before, we calculate the MIMO gain from the measured throughput. We compare our HDA solution with the baseline four-element antenna. The curves are the averaged results from the measurements with the location settings 1, 2, and 3. Figure 6.9 (a) was measured with four enabled STAs. For smaller MCS values up to MCS4, the HDA system and the four-element antenna achieve nearly the same performance. The performance is slightly worse than the performance we measured during the RF cable tests. This might be due to a worse SNR or interference from neighboring Wi-Fi channels or other devices using this frequency band. From MCS5 to MCS8 the four-element antenna can not achieve the same throughput as the HDA system. The throughput is decreasing for higher MCS values since we fix the MCS and the channel does not support MU-MIMO for this MCS. As explained in Section 6.1, the Wi-Fi stack can not select all users due to the ill conditioned channel and drops back to serve the users in time-devision multiple access (TDMA) mode.




(a)




(b)

Figure 6.9.: Mean of the throughput and the MIMO gain for each MCS: (a) with all four STAs; (b) with STAs 1, 2, and 3. It is the mean over all measurements for the location settings 1, 2, and 3.

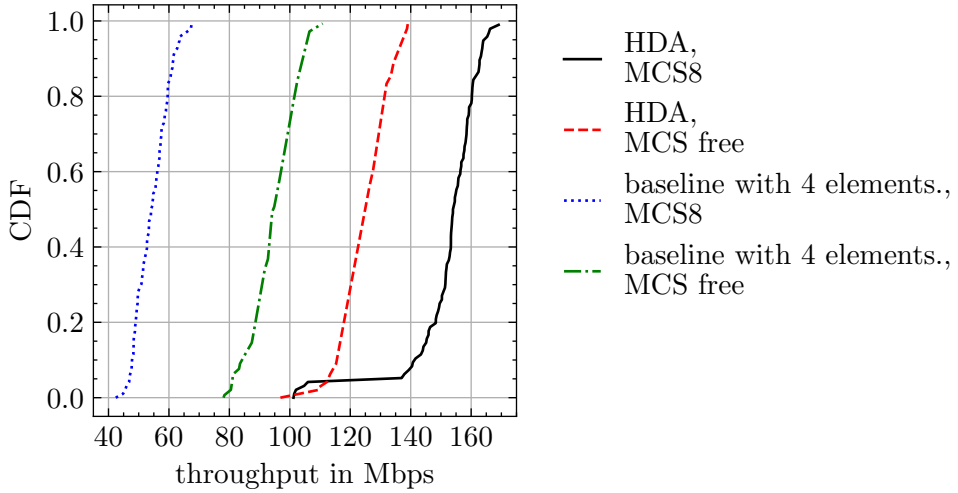



Figure 6.10.: CDF of the throughput of the HDA system and the four-element antenna system for both a fixed MCS of 8 and a free MCS configuration. Combines all measurements with the location settings 1, 2, and 3.

The maximum of the mean of all throughput measurements of the HDA system is 155 Mbps. The maximum of the four-element antenna is 104 Mbps. This is an increase of nearly 50 %.

Figure 6.9 (b) was measured with three STAs, 1, 2, and 3. The results are very similar to the results with four STA. The only difference is that the performance of the baseline system does drop more severely for MCS7 and MCS8. A reason could be, that in the four STA case the MU-MIMO user grouping of the Wi-Fi stack can select three out of four STAs to optimize the MU-MIMO channel. On the contrary to the RF cable measurements, the performance of the HDA system does not decrease because of a fourth STA. The performance is in general slightly worse than the performance of the ideal RF cable setup.

Figure 6.10 is a CDF of the throughput for the fixed MCS8 setting and a free MCS configuration setting. In the free configuration, the Wi-Fi AP NIC can freely choose the MCS index. The CDF is calculated from all measurements with the location settings 1, 2, and 3. The four-element antenna system can not achieve a MU-MIMO gain for MCS8 and, hence, the throughput is very low. For the HDA system, the performance of the free MCS configuration is lower than the enforced MCS8 value. Probably the MU-MIMO user grouping has some security margin when selecting the best MCS index. The average throughput of the HDA system with the free MCS configuration over all measurements is 124 Mbps. The four-element antenna achieves 95 Mbps. This is an increase of 30 %.

Besides the fixed locations, we measured the throughput with a slowly moving STA1 and fixed STAs 2, 3, and 4. STA1 is moving very slowly (i.e., around 6 m in 50 s). The purpose was not to test mobility but to test many locations at the same time. During these

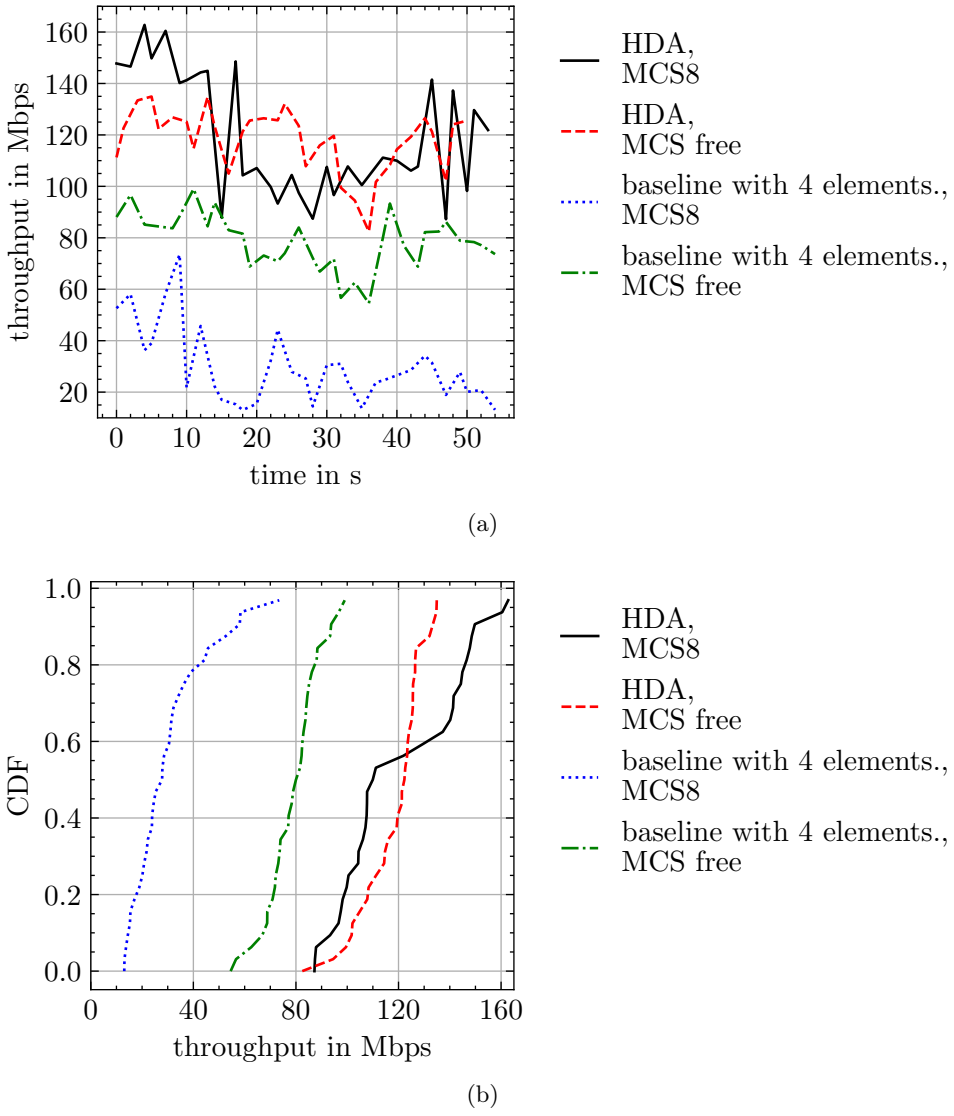



Figure 6.11.: Measurement of the HDA system and the baseline system with a moving STA1 (location setting 4) for a fixed MCS of 8 and a free MCS configuration: (a) The throughput over the measurement time (roughly equivalent to the location); (b) The CDF of the throughput of the whole measurement.

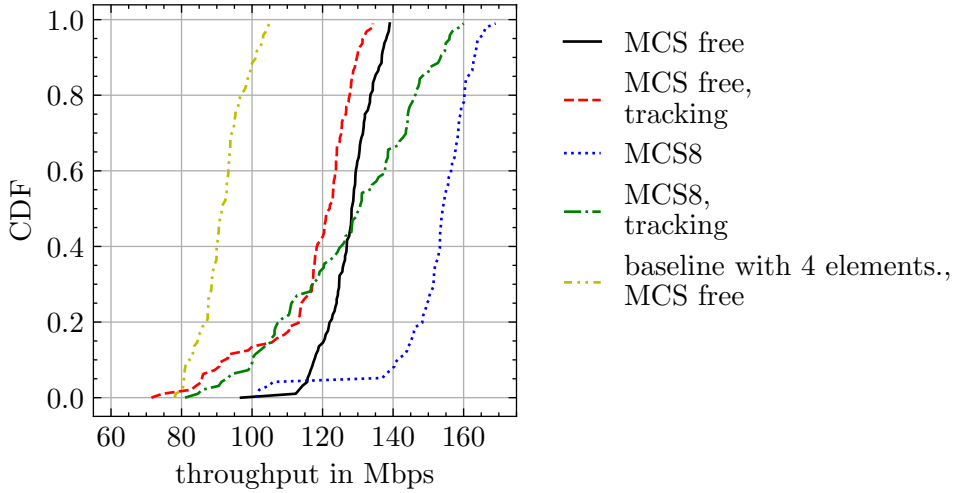



Figure 6.12.: Impact of the beam tracking on the CDF of the throughput of the HDA system. Measurements for a fixed MCS of 8 and a free MCS configuration. The throughput of the four-element antenna system is shown as a baseline. Measured with the location settings 1 and 3.

measurements, the HDA system does use the tracking algorithm. As before, we measured with a fixed MCS8 setting and a free MCS configuration. One can see the measured throughput over time in Figure 6.11 (a) and the CDF of all measurements in Figure 6.11 (b). The four-element antenna can not support MU-MIMO with MCS8 at any location with a decent throughput. The HDA system can achieve a high throughput with MCS8 at some but not all locations. Between 20 s and 40 s, the throughput drops to lower values. During this time STA1 first passed behind a projector and afterwards behind STA3. Hence, first the LOS got blocked and then the inter-user interference increased due to a similar AoA for both stations. The results with the free MCS are very similar to the measurements with the fixed location settings. The mean throughput of the HDA system is 118 Mbps and of the four-element antenna 79 Mbps. This is an increase of nearly 50 %.

### Tracking Impact

As mentioned before, the tracking algorithm is not yet optimized. Nonetheless, we wanted to investigate its impact on the throughput performance. Figure 6.12 shows the CDF of the throughput with and without the tracking enabled for all STAs. The results are combined from multiple measurements with the location settings 1 and 3. The average throughput when the AP NIC can freely select the MCS index without tracking is 127 Mbps and with tracking 117 Mbps. This is a loss of 8 % due to the tracking. The impact on the highest

possible MCS 8 is much larger. The average throughput decreases by 16 % from 152 Mbps to 128 Mbps. The lower beamforming gain of the non-optimal beam direction during the tracking decreases the SINR of the effective channel. In our setup this seems to be enough so that the highest MCS can not be used anymore. In comparison to the four-element antenna, the HDA system with tracking still achieves a better performance.

We noticed during the measurements, that the duration of the tracking phase for the MCS8 setup was much longer than for the free MCS configuration. The duration in our implementation is variable and determined by the successful measurement of a packet in the uplink (which is triggered by the `ping` command). This might be an effect of the Wi-Fi user grouping in the driver/firmware of the NIC, especially with the SINR close to the limit for the fixed MCS. Of course, the longer the beamforming points towards the non-optimal direction the lower the average performance will be. If the tracking could be integrated in the driver (e.g., using MCS0 and a high priority for the tracking packets) the duration and the impact of the tracking algorithm might be reduced.

## 6.5. Summary

We presented a concept to use hybrid digital-analog (HDA) beamforming for Wi-Fi networks. Typical commodity Wi-Fi systems do not achieve the full MU-MIMO capabilities of the current IEEE 802.11ac/ax standard. Our concept does increase the MU-MIMO performance. The concept does not increase the complexity like a full digital massive MIMO approach and, more importantly, it is based on COTS Wi-Fi systems. It does not require any change to the Wi-Fi protocol.

We implemented the concept and build a full demonstrator. The demonstrator is based on our self-developed analog beamforming module and commodity Wi-Fi hardware. Furthermore, we investigated the required algorithms and developed a control protocol for the HDA system. We evaluated the concept and the demonstrator. First, we proved by measuring the condition number of the effective channel that the HDA concept does increase the channel orthogonality as suggested in the introduction. Second, we measured key aspects and the end-to-end throughput performance of the demonstrator. The demonstrator is limited by the COTS Wi-Fi component to a maximum of three MU-MIMO streams. The HDA system could achieve this limit in nearly all measurements. In contrast, a four-element antenna system could not achieve the limit for all MCSs. The performance of the HDA system is 50 % higher than the baseline system.

# 7. Conclusion and Outlook

The MU-MIMO technology is an important cornerstone for the improvement of multiple wireless communication standards. It is one of the key performance drivers for 5G. Also, for WLANs, it boosts the throughput of the current IEEE 802.11 standard. The performance increase can be tremendous, especially, when a massive number of antenna elements is used. Unfortunately, the implementation of such massive MIMO systems exhibits various challenges, from complexity issues to the problem of the channel estimation. A system can be infeasible if the hardware technology does not allow the complexity of the design. The hybrid digital-analog (HDA) architecture is a solution to these problems. It reduces the complexity of systems with a large number of antenna elements compared to the standard fully-digital architecture. Nearly all research on mm-wave systems for 5G is based on the hybrid approach for massive MIMO. This work explained the advantages, the challenges, and the implementation of the HDA architecture. We derived application scenarios for the hybrid architecture from its properties. The main application scenario is the usage in complexity-limited MU-MIMO systems, like, mm-wave massive MIMO systems. The hybrid concept can also be applied to reduce the channel estimation overhead similar to the JSDM approach. Last but not least, a hybrid beamforming extension can improve the MU-MIMO performance of an existing system without the need to change the communication protocol.

Many works in the literature only cover single aspects of the HDA approach. Often, the hardware design and the signal processing are independently investigated. In this thesis, we tried to investigate hybrid systems holistically. We proposed selected algorithms from the literature for the main signal processing problems related to the hybrid architecture. The algorithm for the initial beam alignment (BA) between the BS/AP and the UEs achieves high performance. It is robust against fast channel variations, can be used with different hardware structures, and works with realistic hardware assumptions. The selected hybrid MU-MIMO precoding algorithm decouples the analog beamforming from the digital precoding. The analog beamforming points standard Fourier-based single-beam patterns towards the scheduled users. The digital precoder based on zero-forcing can further increase the data rate by removing the residual interference. The beamforming creates an effective channel that is more diagonally dominant for the digital precoding. Thus, the power penalty of the digital precoder is reduced, and the combined data rate is optimized. Following the

introduction of the signal processing algorithms, we have investigated the impact of multiple hardware aspects on the system performance. We considered both the structure of the analog beamforming network and its components. Through simulations, we could determine the best parameters of the hardware to improve the signal processing performance. We compared the two main structures of the analog network. The OSPS structure achieves a better system performance compared to the FC structure. the OSPS structure has lower complexity and higher power efficiency than the FC structure. Besides the structure, we also analyzed the effect of the bit width of the vector modulators on the system performance. We showed that already 4 bits are sufficient for 64 and even 256 antenna elements. In addition, we covered other aspects of the hardware of the analog network, e.g., the power distribution network and the amplification design.

To enable a real-life evaluation of the hybrid concept, we developed a hybrid testbed for the frequency range between 2.2 GHz and 2.5 GHz. The core part of the testbed is an analog beamforming module. It is an analog network with 2 RF chain ports and 16 antenna ports. The phase and the amplitude can be set with a resolution of 8 bits. Multiple modules can be interconnected to allow for a larger number of RF chains or antenna elements. We describe the hardware design of the analog module and the calibration procedure to achieve the high resolution.

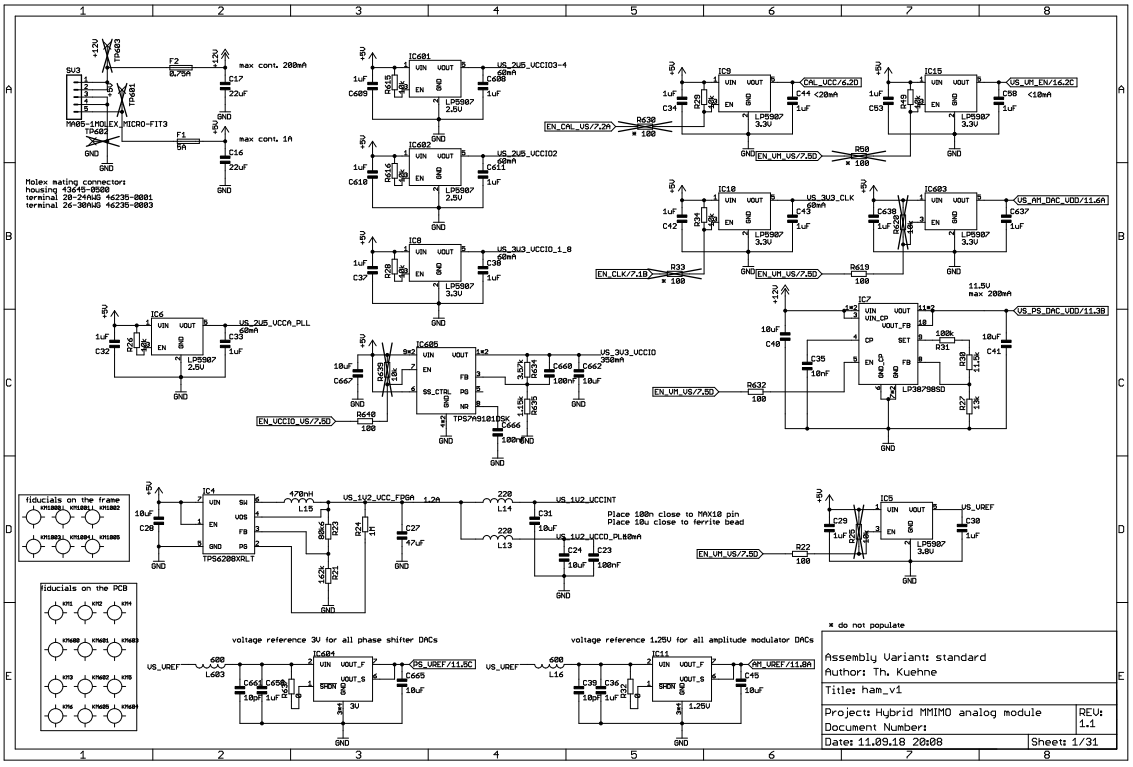
The second part of the thesis focuses on two application scenarios for the HDA architecture. We propose a system design for each of the two scenarios. The system designs are based on the signal processing results and hardware aspect analysis of the first part of the thesis. The first system is a BS for a 5G mm-wave small cell. The hardware and the HDA beamforming system were developed in the European project SERENA. We described the system specifications, the underlying signal processing algorithms, and the frame format based on the 5G standard. In a simulation, we present the performance of the system. The simulation includes the antenna characteristic, hardware models, and the signal processing algorithms. We used 5G physical layer frame parameters and the QuaDriGa 3-D channel simulator with 3GPP mm-wave channel parameters. In the simulated small cell scenario, the system shows a high BA performance and a large MU-MIMO gain. The second system was developed to demonstrate the application of the hybrid beamforming concept for Wi-Fi networks. The concept is based on COTS Wi-Fi hardware and does increase the MU-MIMO performance by using 32 antenna elements. The concept does not increase the complexity like the fully digital massive MIMO approach. It does not require any change to the Wi-Fi protocol. We implemented the concept, the signal processing, and a control protocol for the hybrid beamforming. We build a full operational demonstrator. The demonstrator is based on the self-developed analog beamforming module for the 2.2 GHz to 2.5 GHz frequency range and commodity Wi-Fi hardware. We evaluated the concept by measuring

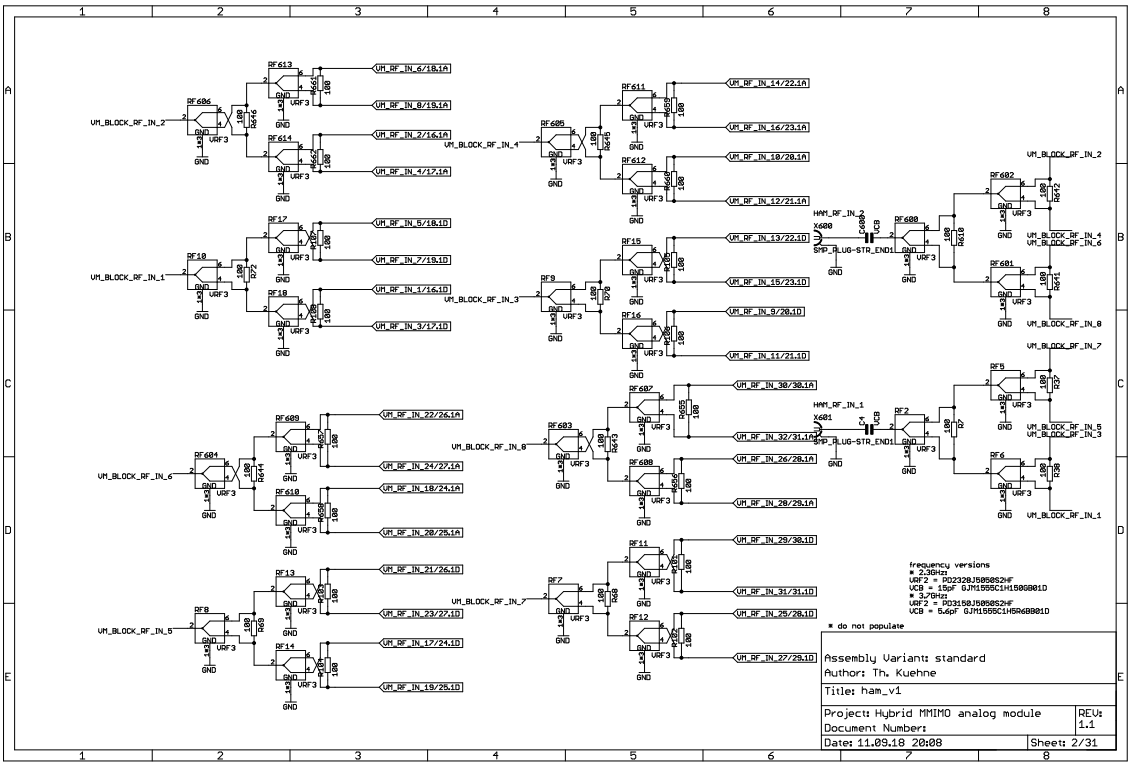
different aspects and benchmarks of the hybrid approach. The HDA system could achieve the MU-MIMO limit of the Wi-Fi hardware in nearly all measurements. In contrast, a four-element antenna baseline system could not achieve the limit for all modulation schemes. The performance of the HDA system is 50 % higher than the baseline system.

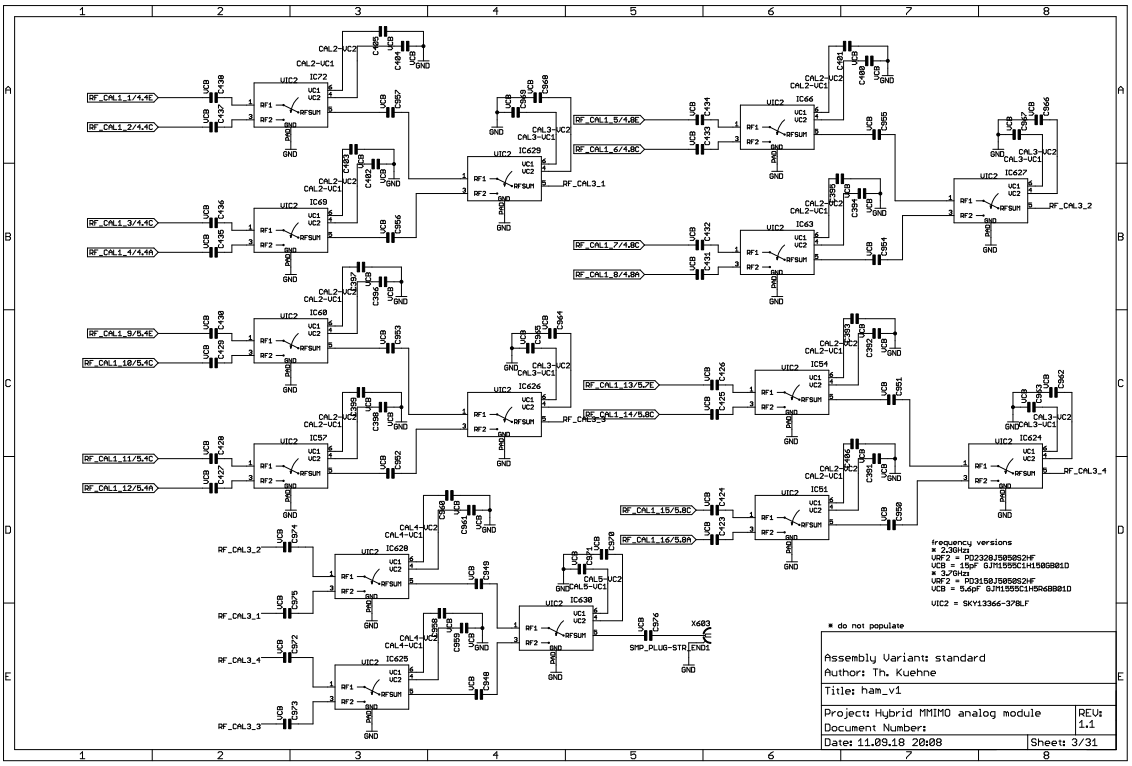
In conclusion, we have investigated various aspects of the HDA architecture including the interconnection of the system design, signal processing, and hardware implementation. We could prove the performance increase with simulations and by measuring the throughput of a self-developed testbed. Nonetheless, there are different open research directions that could follow up on the presented work. One apparent next step for the mm-wave application is the realization of a testbed. Unfortunately, the proposed system for the European project SERENA was not finished by the time this thesis was written. The HDA architecture will remain in the focus of mobile communication research. Future systems for the sixth generation standard for cellular networks are supposed to operate in even higher frequencies above the 5G mm-wave bands. The bandwidth of these systems will be larger, and the hardware complexity will remain a challenging problem. The HDA architecture can help realize these systems. The assumptions for the mm-wave channel need to be reviewed for the higher frequencies in the Terahertz bands. Due to the large bandwidth, the narrowband assumption is debatable and true time delay concepts might become important. For the signal processing, we have only briefly covered the problem of beam tracking. If the user is moving, also the second-order statics (i.e., the AoA/AoD) slowly change. The change can either be estimated by a reoccurring BA procedure or by a beam tracking algorithm. There are first works in the literature, which propose sophisticated tracking algorithms with improved performance. Of course, the much-discussed machine learning is a candidate to revolutionize the signal processing. We can imagine a holistic signal processing where the problems of the BA, the hardware calibration, and the beamforming pattern calculation are solved by reinforcement learning or other modern machine learning algorithms.

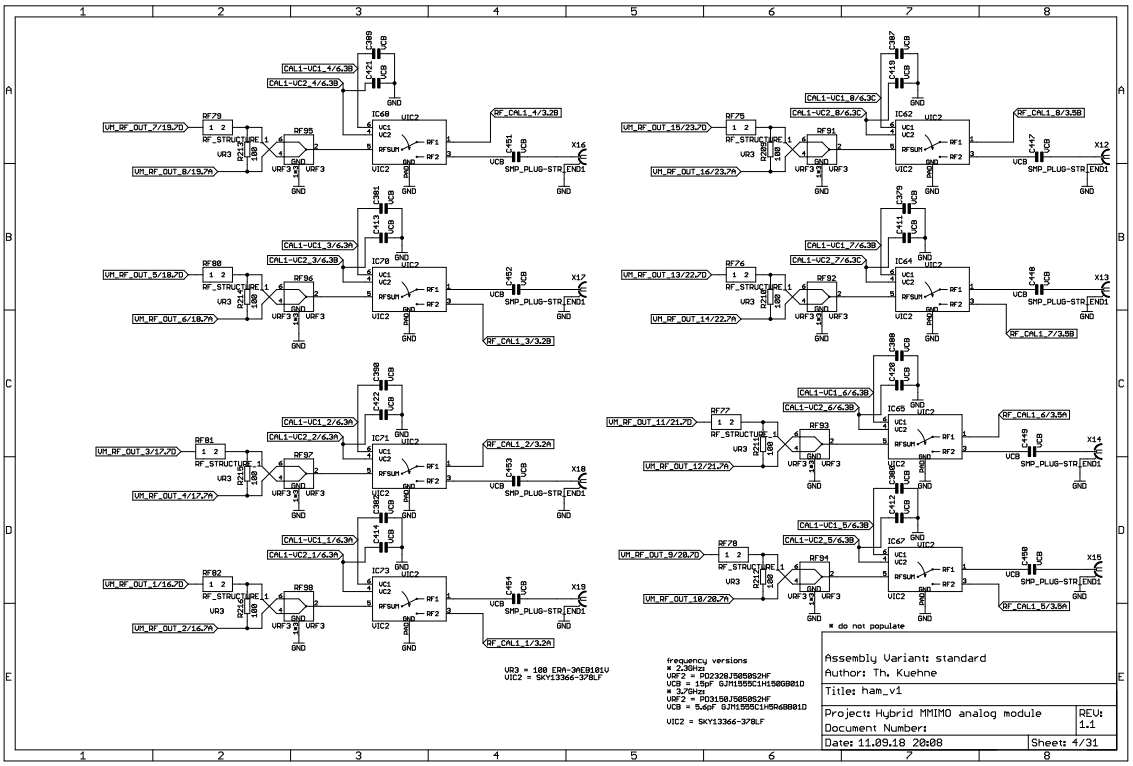
# Appendix A.

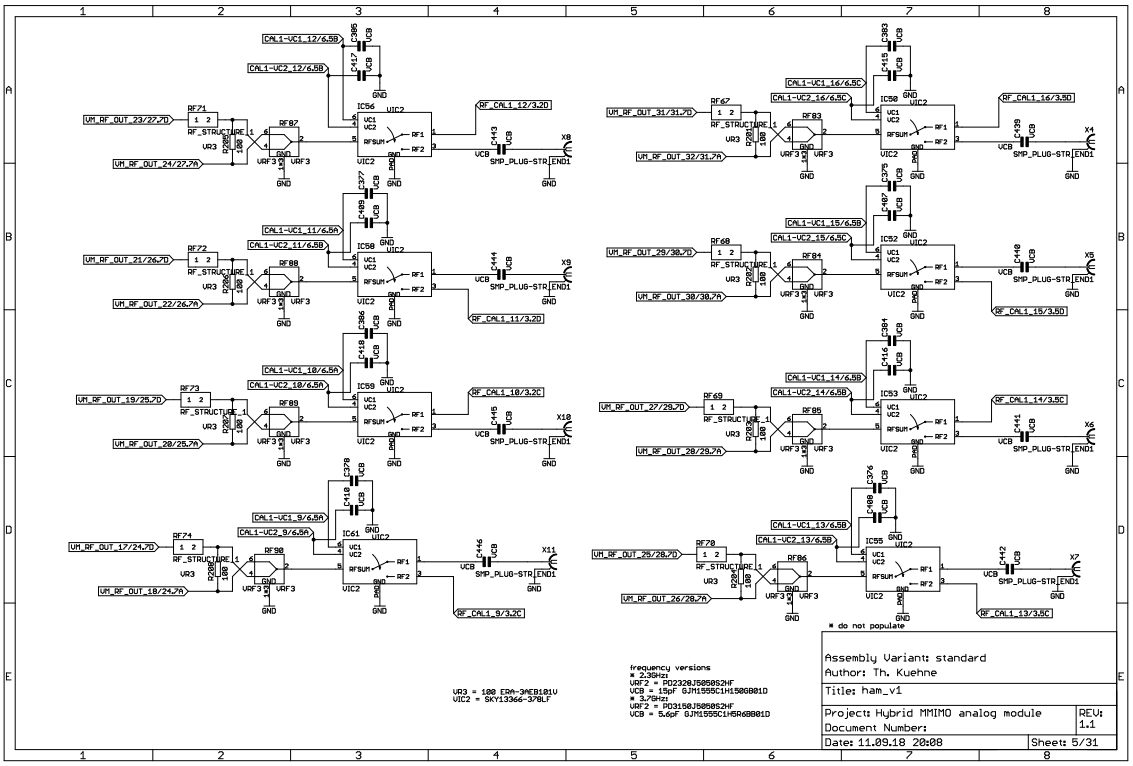
# Schematic of the Analog Module from Chapter 4

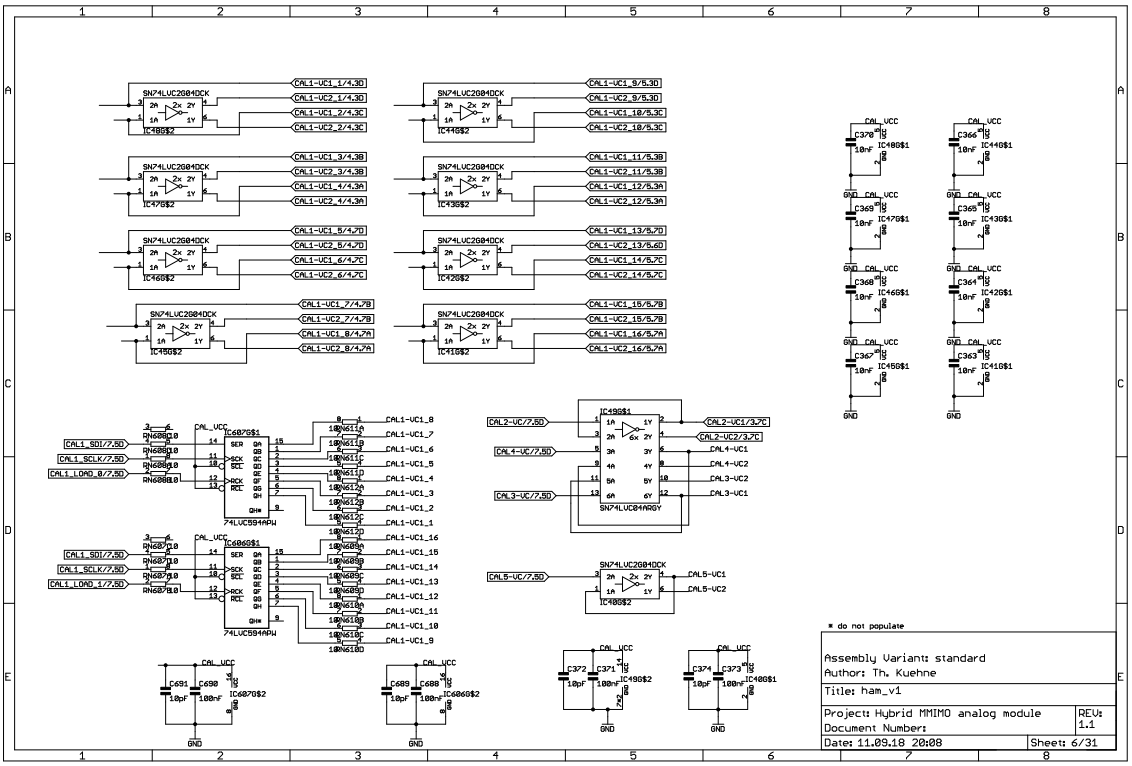


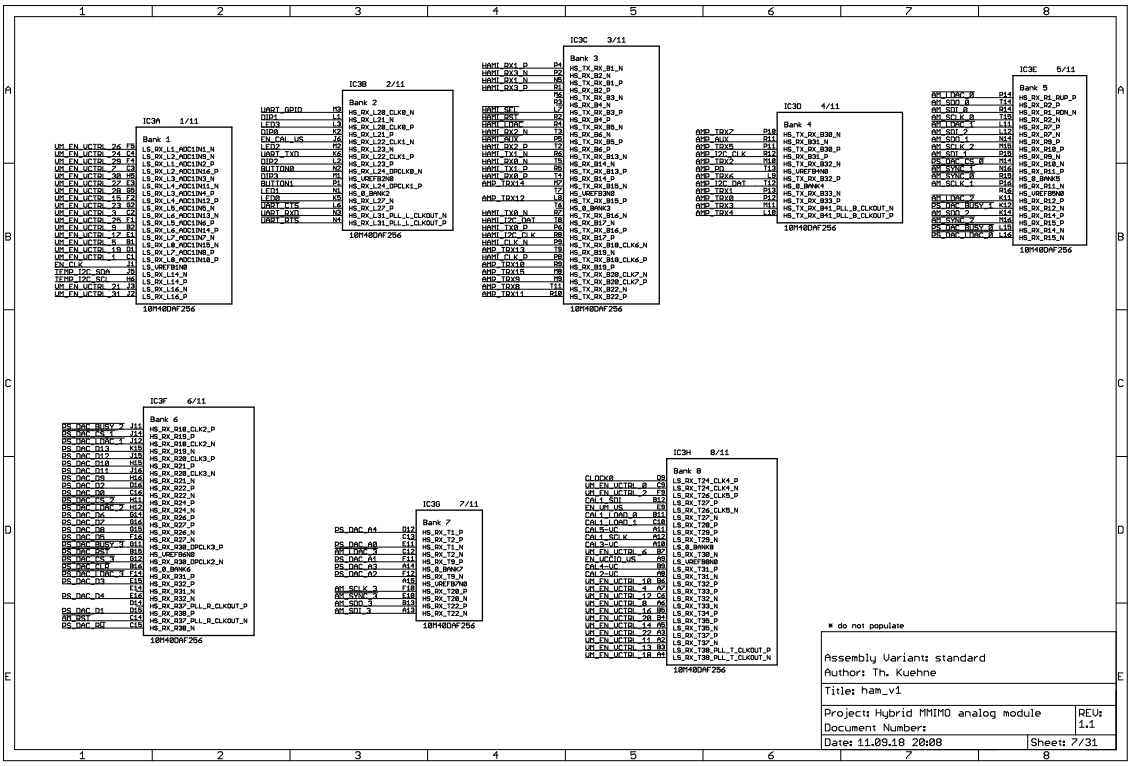

frequency versions
* 2.3GHz
VAF2 = PD2320J5050S2HF
VCB = 15pF GJM1555C1H150GB01D
* 3.7GHz
VAF2 = PD3150J5050S2HF
VCB = 5.6pF GJM1555C1H5R6BB01D
* do not populate
Assembly Variant: standard
Author: Th. Kuehne
Title: ham_v1
Project: Hybrid MMIMO analog module
Document Number:
REV: 1.1
Date: 11.09.18 20:08
Sheet: 2/31

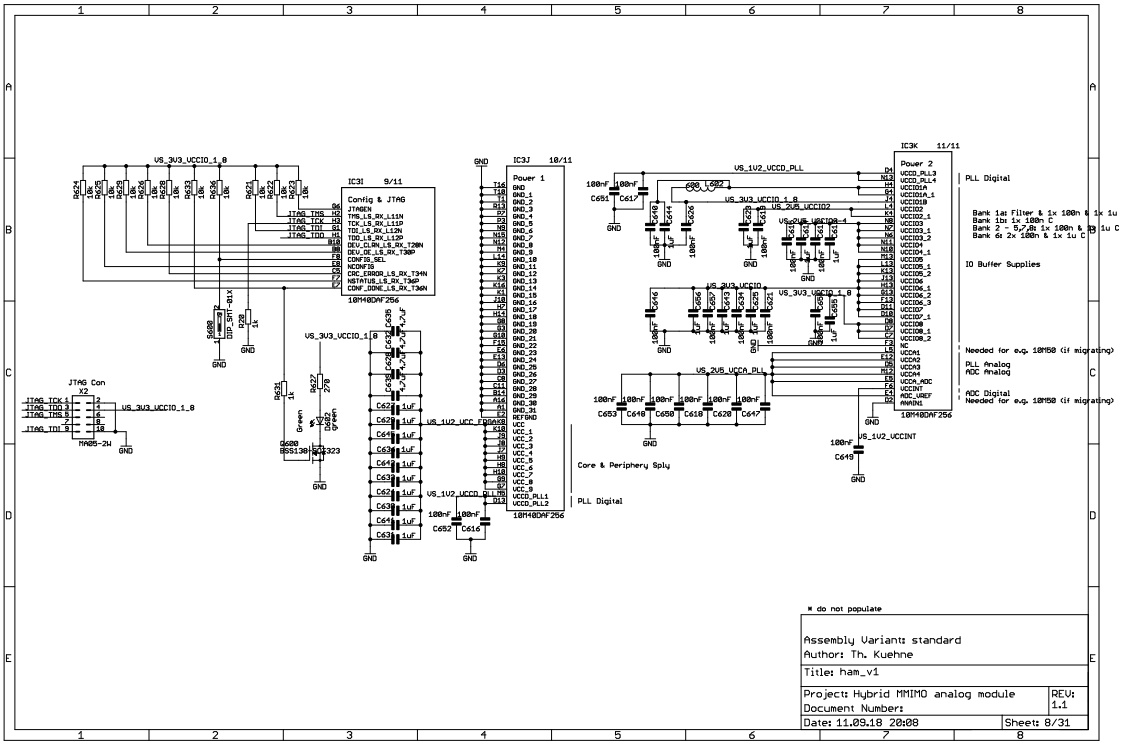

frequency versions
* 2.3GHz
VAF2 = PD2320J5050S2HF
VCB = 15pF GJM1555C1H150GB01D
* 3.7GHz
VAF2 = PD3150J5050S2HF
VCB = 5.6pF GJM1555C1H5R6BB01D
VIC2 = SKY13366-378LF
* do not populate
Assembly Variant: standard
Author: Th. Kuehne
Title: ham_v1
Project: Hybrid MMIMO analog module
Document Number:
REV: 1.1
Date: 11.09.18 20:08
Sheet: 3/31

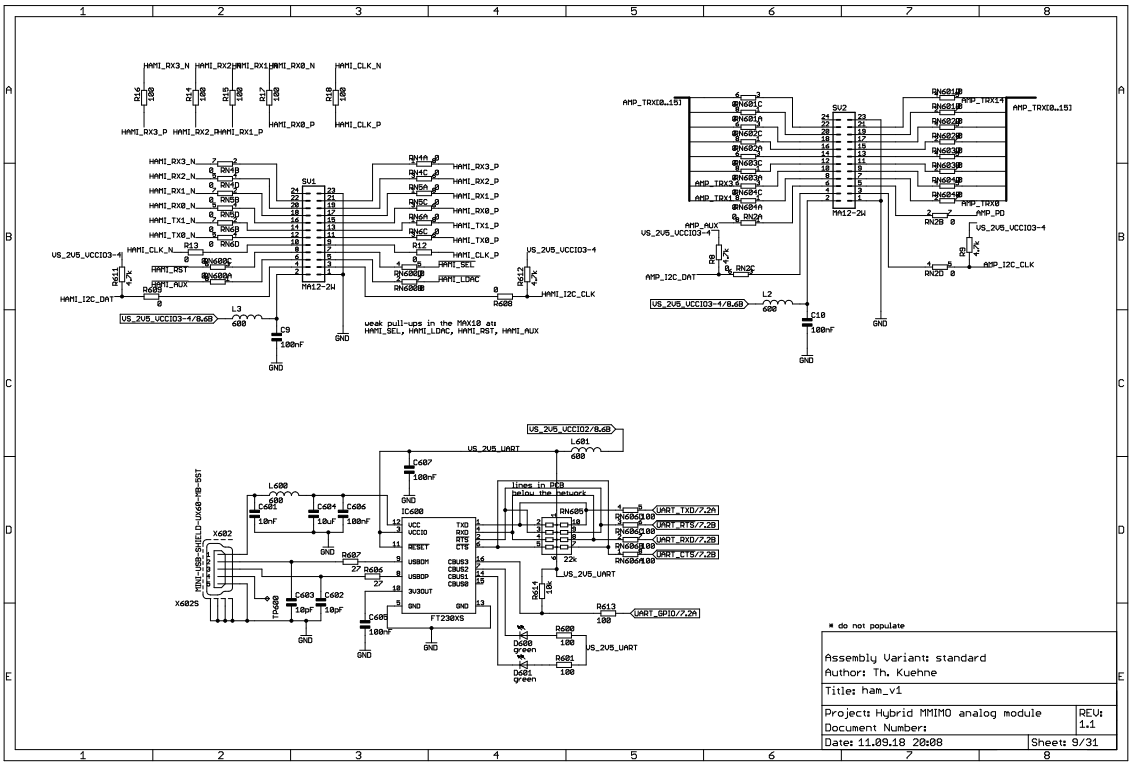

VR3 = 100 ERA-3AEB101V
VIC2 = SKY13366-378LF
frequency versions
= 2.3GHz
VRF2 = PD2328J5050S2HF
VCB = 15pF GJM1555C1H150GB01D
= 3.7GHz
VRF2 = PD3150J5050S2HF
VCB = 5.6pF GJM1555C1H5R6BB01D
VIC2 = SKY13366-378LF
* do not populate
Assembly Variant: standard
Author: Th. Kuehne
Title: ham_v1
Project: Hybrid MMIMO analog module
Document Number:
REV: 1.1
Date: 11.09.18 20:08
Sheet: 4/31

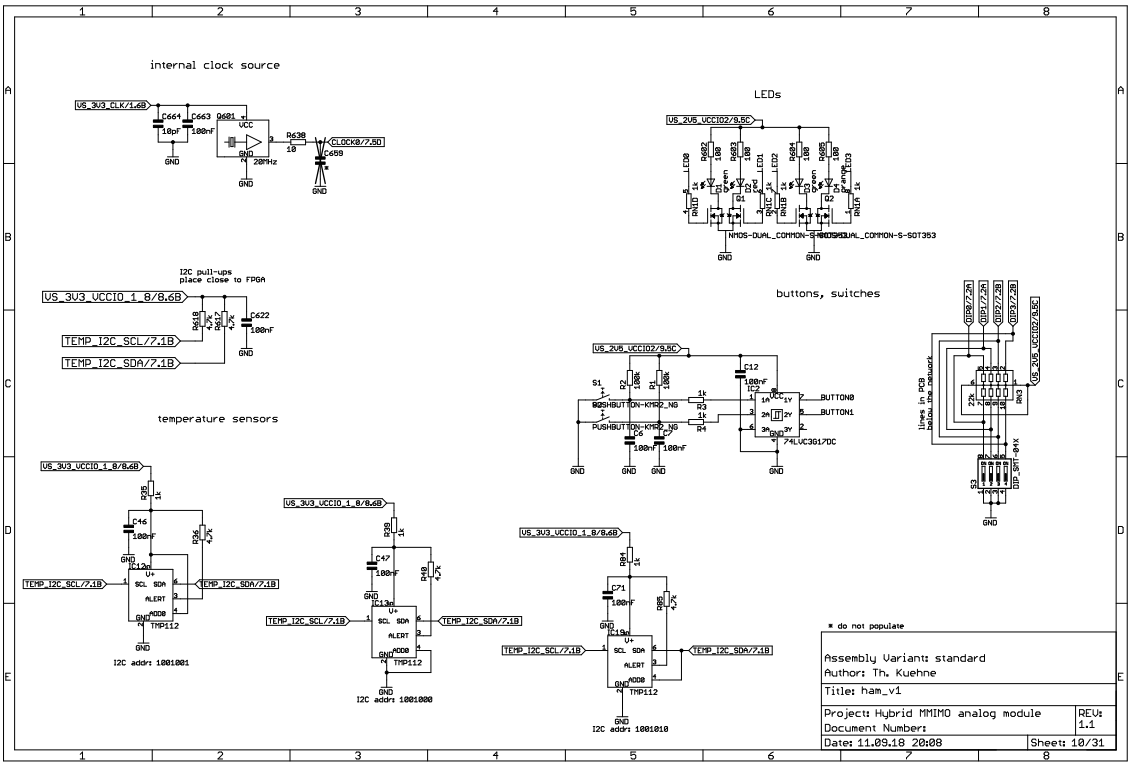

VR3 = 100 ERA-3AEB101V
VIC2 = SKY13366-378LF
frequency versions
= 2.3GHz
VRF2 = PD2328J5050S2HF
VCB = 15pF GJM1555C1H150GB01D
= 3.7GHz
VRF2 = PD3150J5050S2HF
VCB = 5.6pF GJM1555C1H5R6BB01D
* do not populate
Assembly Variant: standard
Author: Th. Kuehne
Title: ham_v1
Project: Hybrid MMIMO analog module
Document Number:
REV: 1.1
Date: 11.09.18 20:08
Sheet: 5/31

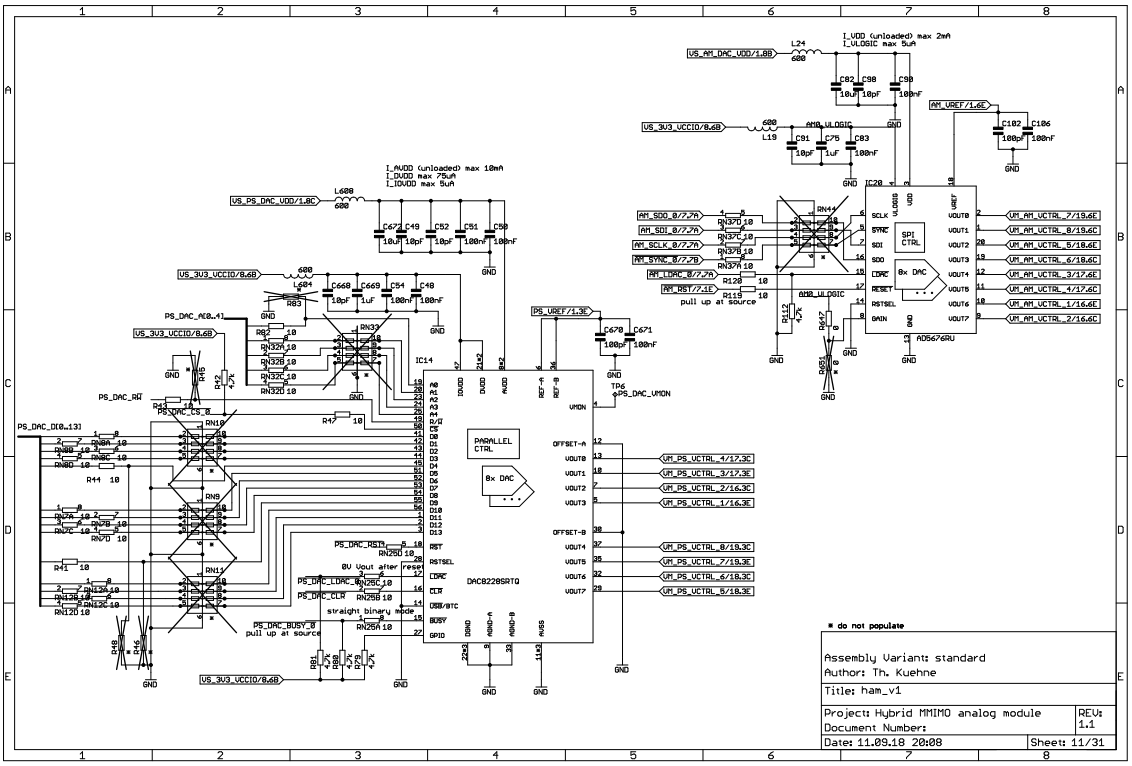

* do not populate
Assembly Variant: standard
Author: Th. Kuehne
Title: ham_v1
Project: Hybrid MMIMO analog module
Document Number:
REV: 1.1
Date: 11.09.18 20:08
Sheet: 6/31

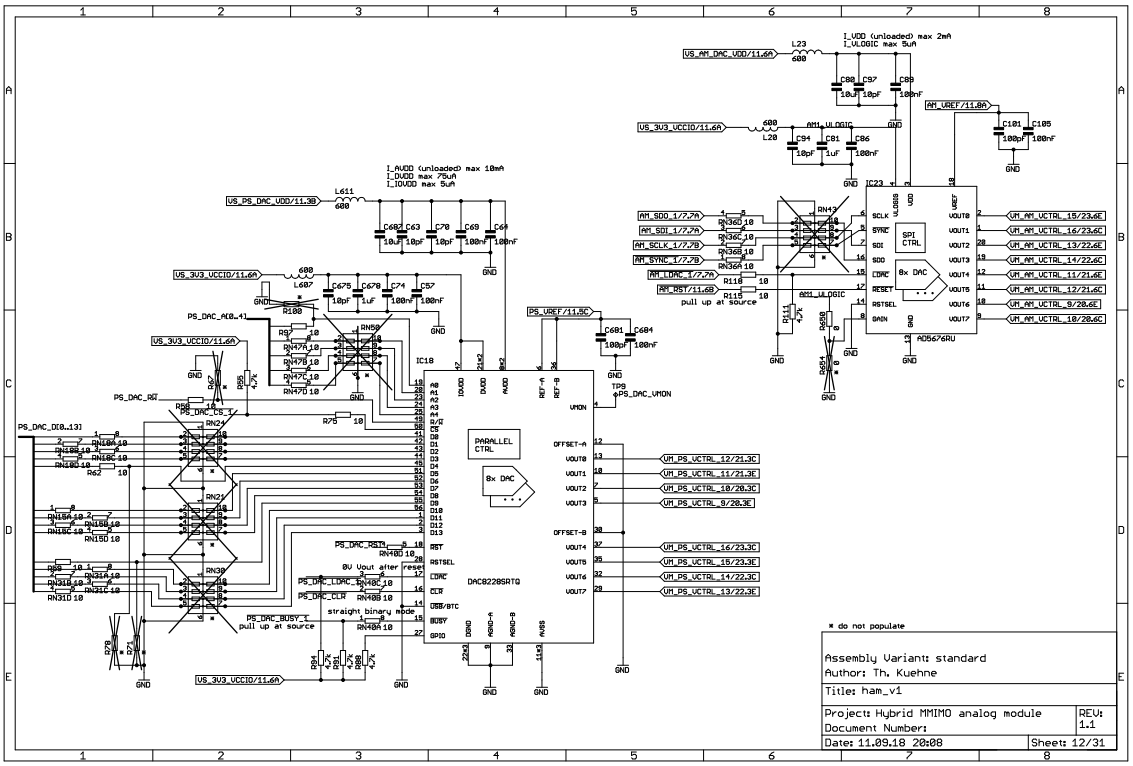

* do not populate
Assembly Variant: standard
Author: Th. Kuehne
Title: ham_v1
Project: Hybrid MMIMO analog module
Document Number:
REV: 1.1
Date: 11.09.18 20:08
Sheet: 7/31

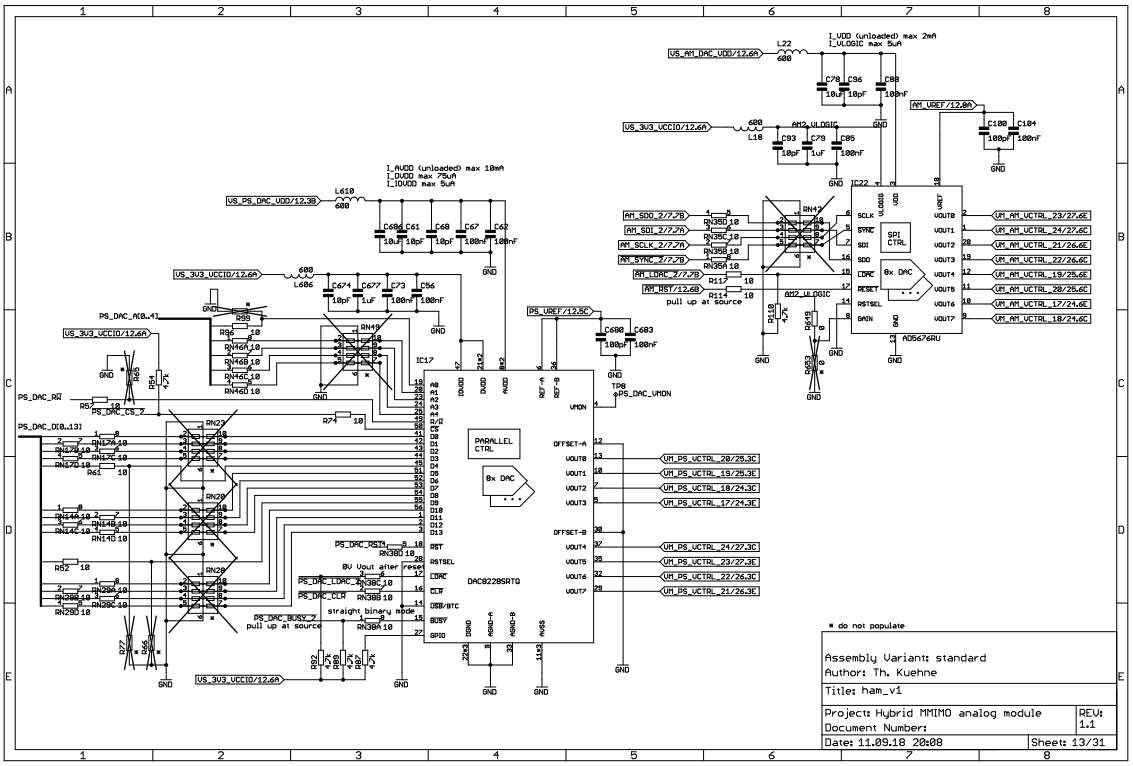

IC3I 9/11
Config & JTAG
10M40DAF256
IC3J 10/11
Power 1
IC3K 11/11
Power 2
PLL Digital
IO Buffer Supplies
Core & Periphery Sply
JTAG Con
X2
MA05-2W
* do not populate
Assembly Variant: standard
Author: Th. Kuehne
Title: ham_v1
Project: Hybrid MMIMO analog module
Document Number:
REV: 1.1
Date: 11.09.18 20:08
Sheet: 8/31

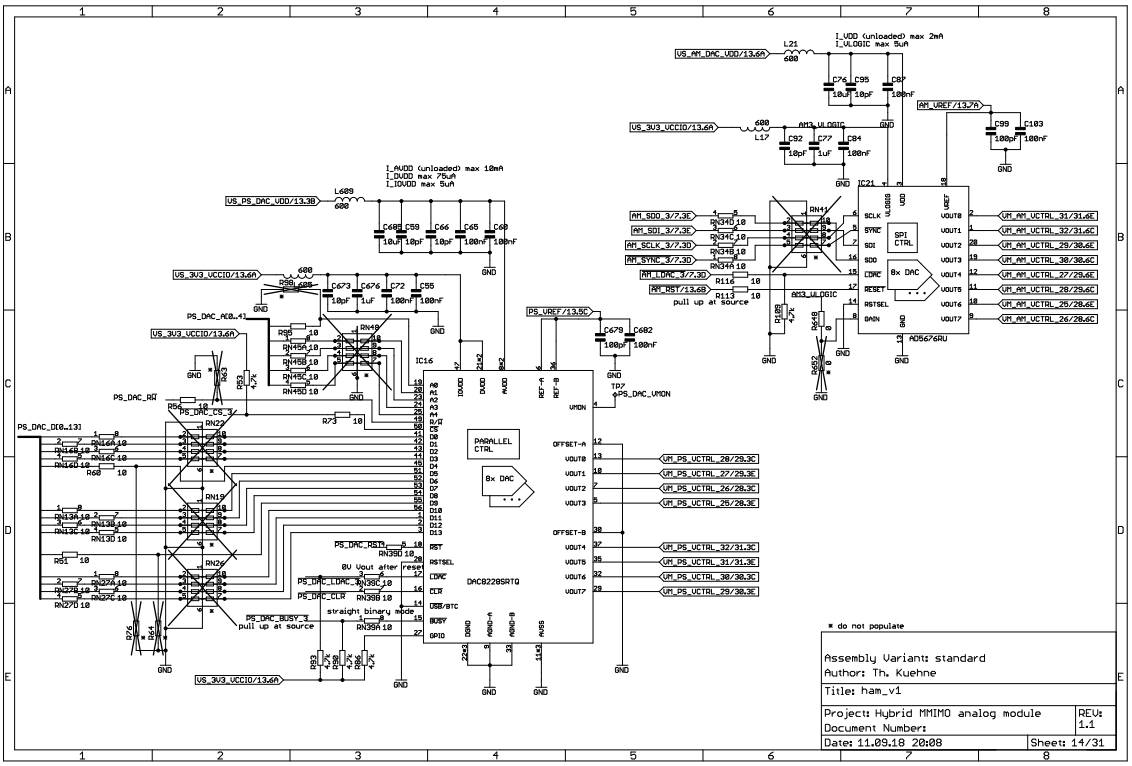

SV1
SV2
MA12-2W
L3
C9
100nF
weak pull-ups in the MAX10 at HAMI_SEL, HAMI_LDAC, HAMI_RST, HAMI_AUX
L2
C10
100nF
X602
MINI-USB-SHIELD-UX60-MB-5ST
FT230XS
IC600
L600
L601
C687
100nF
lines in PCB below the network
UART_TXD/7,2A
UART_RTS/7,2A
UART_RXD/7,2A
UART_CTS/7,2A
UART_GPIO/7,2A
* do not populate
Assembly Variant: standard
Author: Th. Kuehne
Title: ham_v1
Project: Hybrid MMIMO analog module
Document Number:
REV: 1.1
Date: 11.09.18 20:08
Sheet: 9/31

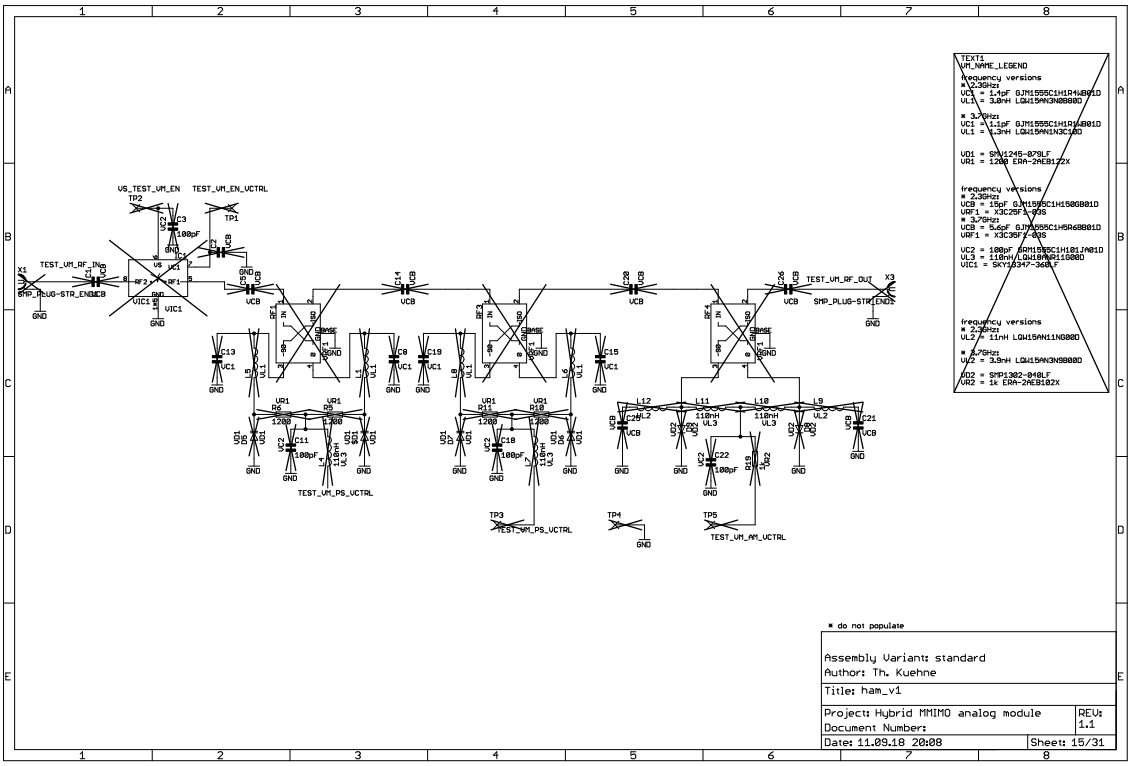

internal clock source
LEDs
I2C pull-ups
place close to FPGA
buttons, switches
temperature sensors
I2C addr: 1001001
I2C addr: 1001000
I2C addr: 1001010
* do not populate
Assembly Variant: standard
Author: Th. Kuehne
Title: ham_v1
Project: Hybrid MMIMO analog module
Document Number:
REV: 1.1
Date: 11.09.18 20:08
Sheet: 10/31

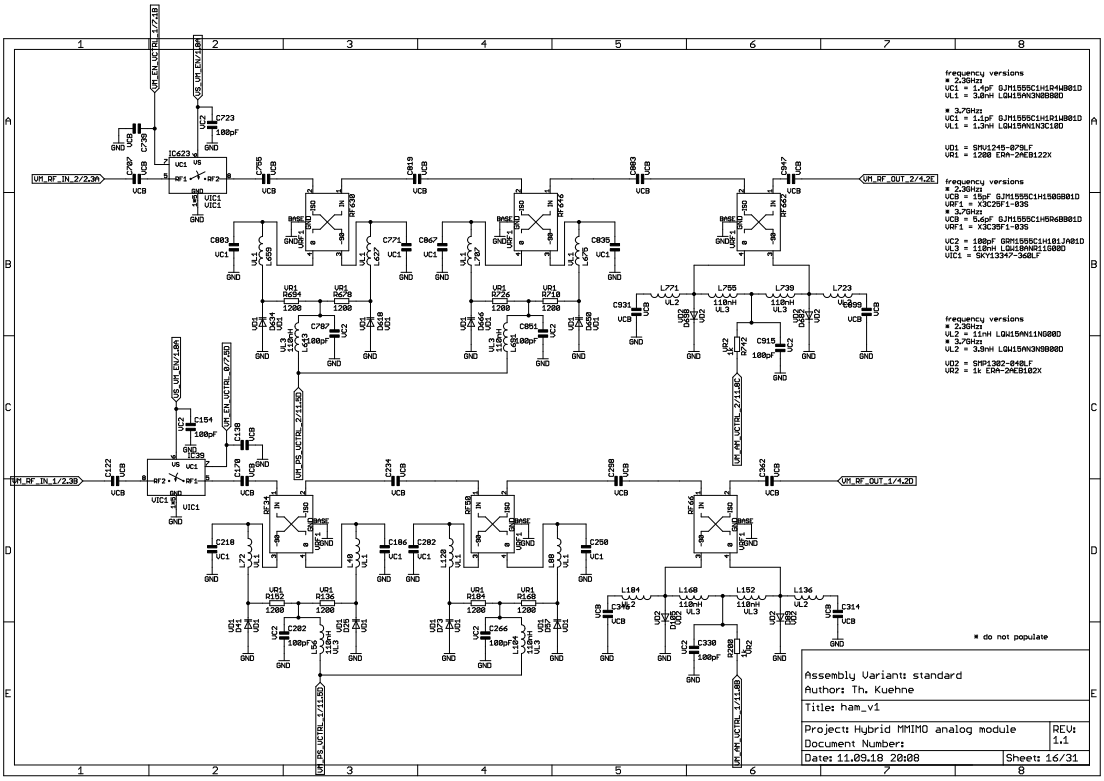

straight binary mode
pull up at source
pull up at source
* do not populate
Assembly Variant: standard
Author: Th. Kuehne
Title: ham_v1
Project: Hybrid MMIMO analog module
Document Number:
REV: 1.1
Date: 11.09.18 20:08
Sheet: 11/31

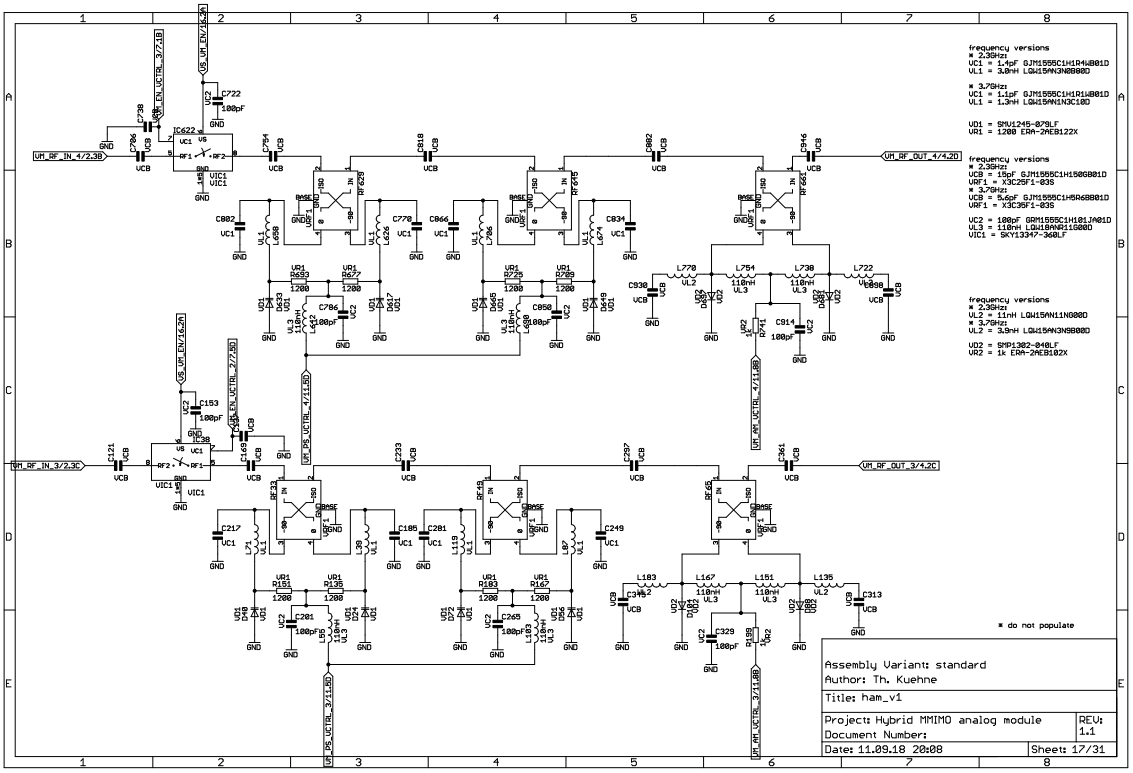

I_VDD (unloaded) max 2mA
I_VLOGIC max 5uA
VS_AM_DAC_VDD/11.6A
L23
600
C98 10uF
C97 10pF
C89 100nF
AM_VREF/11.6A
VS_3V3_VCCIO/11.6A
L20
AM1_VLOGIC
C94 10pF
C81 1uF
C86 100nF
C101 100pF
C105 100nF
GND
IC23
SCLK
SYNC
SDI
SDO
CLRC
RESET
RSTSEL
SMIN
VLOGIC
VDD
VREF
SPI CTRL
8x DAC
VOUT0
VOUT1
VOUT2
VOUT3
VOUT4
VOUT5
VOUT6
VOUT7
GND
AD5676RU
AM_SDO_1/7.7A
AM_SDI_1/7.7A
AM_SCLK_1/7.7B
AM_SYNC_1/7.7B
AM_LDAC_1/7.7A
AM_RST/11.6B
pull up at source
RN47
RN36C 10
RN36B 10
RN36A 10
R116 10
R111 4.7K
R108 4.7K
AM_VM_UCTRL_15/23.6E
AM_VM_UCTRL_16/23.6C
AM_VM_UCTRL_13/22.6E
AM_VM_UCTRL_14/22.6C
AM_VM_UCTRL_11/21.6E
AM_VM_UCTRL_12/21.6C
AM_VM_UCTRL_9/20.6E
AM_VM_UCTRL_10/20.6C
I_AVDD (unloaded) max 10mA
I_DVDD max 75uA
I_IOVDD max 5uA
VS_PS_DAC_VDD/11.3B
L611
600
C662 10uF
C63 10pF
C70 100nF
C69 100nF
C64 100nF
VS_3V3_VCCIO/11.6A
L607
C675 10pF
C678 1uF
C74 100nF
C57 100nF
R100
R97 10
PS_DAC_A[0..4]
RN49
RN47A 10
RN47B 10
RN47C 10
RN47D 10
R95 4.7K
R96 10
PS_DAC_RST
PS_DAC_CS_1
R75 10
PS_VREF/11.5C
C681 100pF
C684 100nF
TP8
PS_DAC_VMON
IC18
AVDD
DVDD
IOVDD
REF-A
REF-B
VMON
PS_DAC_D[0..13]
RN24
RN21
RN39
RN18A 10
RN18B 10
RN18C 10
RN18D 10
RN15A 10
RN15B 10
RN15C 10
RN15D 10
RN31A 10
RN31B 10
RN31C 10
RN31D 10
R62 10
R58 10
PARALLEL CTRL
8x DAC
OFFSET-A
OFFSET-B
DAC8228SRTQ
RST
RSTSEL
CDAC
CLR
OSB/BTC
BUSY
GPIO
DGND
AGND-A
AGND-B
AVSS
PS_DAC_RST
0V Vout after reset
PS_DAC_LDAC
PS_DAC_CLR
straight binary mode
PS_DAC_BUSY_1
pull up at source
RN48A 10
RN48B 10
RN48C 10
RN48D 10
R79
R71
R94 4.7K
R91 4.7K
R88 4.7K
VM_PS_UCTRL_12/21.3C
VM_PS_UCTRL_11/21.3E
VM_PS_UCTRL_10/20.3C
VM_PS_UCTRL_9/20.3E
VM_PS_UCTRL_16/23.3C
VM_PS_UCTRL_15/23.3E
VM_PS_UCTRL_14/22.3C
VM_PS_UCTRL_13/22.3E
2243
33
1143
do not populate
Assembly Variant: standard
Author: Th. Kuehne
Title: ham_v1
Project: Hybrid MMIMO analog module
Document Number:
REV: 1.1
Date: 11.09.18 20:08
Sheet: 12/31

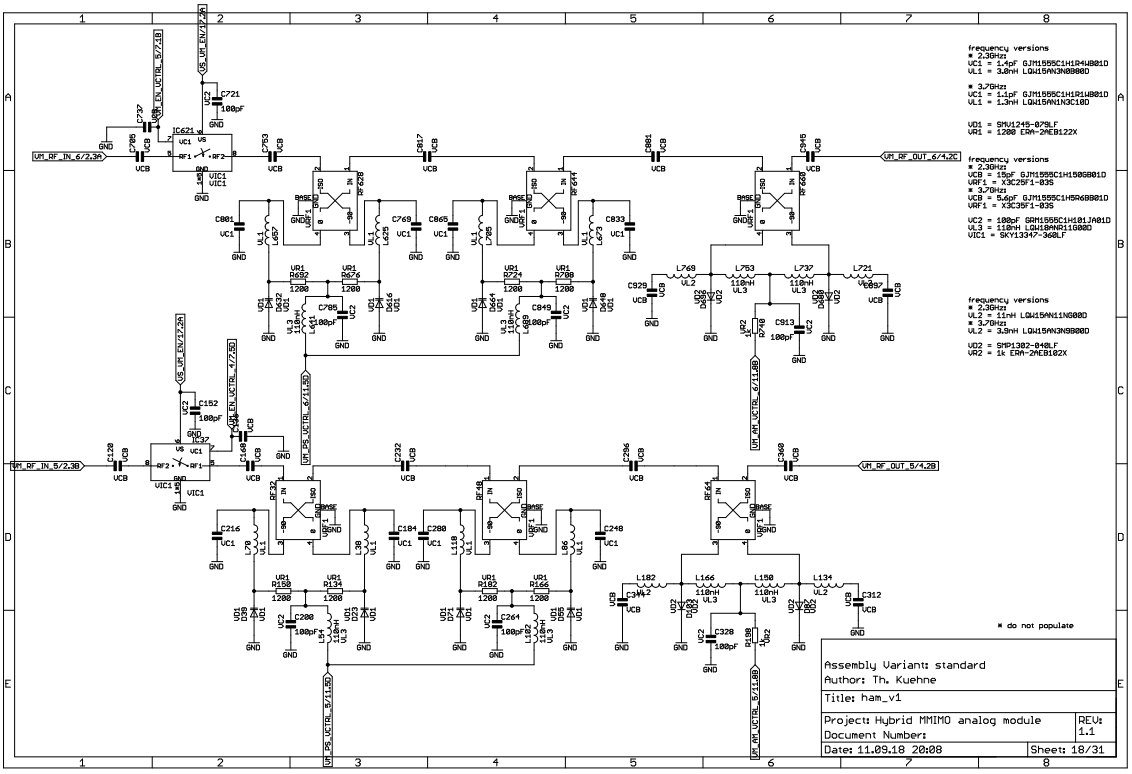

I_VDD (unloaded) max 2mA
I_VLOGIC max 5uA
VS_AM_DAC_VDD/12.6A
L22
600
C78 10uF
C96 10pF
C88 100nF
AM_VREF/12.8A
VS_3V3_VCCIO/12.6A
L18
AM2_VLOGIC
C93 10pF
C79 1uF
C85 100nF
C100 100pF
C104 100nF
GND
IC22
SCLK
SYNC
SDI
SDO
CLRC
RESET
RSTSEL
SMIN
VLOGIC
VDD
VREF
SPI CTRL
8x DAC
VOUT0
VOUT1
VOUT2
VOUT3
VOUT4
VOUT5
VOUT6
VOUT7
AD5676RU
AM_SDO_2/7.7B
AM_SDI_2/7.7A
AM_SCLK_2/7.7A
AM_SYNC_2/7.7B
AM_LDAC_2/7.7B
AM_RST/12.6B
pull up at source
RN47
RN36C 10
RN36B 10
RN36A 10
R117 10
R114 10
R110 4.7K
R105 4.7K
AM_VM_UCTRL_23/27.6E
AM_VM_UCTRL_24/27.6C
AM_VM_UCTRL_21/26.6E
AM_VM_UCTRL_22/26.6C
AM_VM_UCTRL_19/25.6E
AM_VM_UCTRL_20/25.6C
AM_VM_UCTRL_17/24.6E
AM_VM_UCTRL_18/24.6C
I_AVDD (unloaded) max 10mA
I_DVDD max 75uA
I_IOVDD max 5uA
VS_PS_DAC_VDD/12.3B
L610
600
C682 10uF
C61 10pF
C68 100nF
C67 100nF
C62 100nF
VS_3V3_VCCIO/12.6A
L606
C674 10pF
C677 1uF
C73 100nF
C56 100nF
R99
R96 10
PS_DAC_A[0..4]
RN49
RN46A 10
RN46B 10
RN46C 10
RN46D 10
R95 4.7K
R94 4.7K
R97 10
PS_DAC_RST
PS_DAC_CS_2
R74 10
PS_VREF/12.5C
C680 100pF
C683 100nF
TP8
PS_DAC_VMON
IC17
AVDD
DVDD
IOVDD
REF-A
REF-B
VMON
PS_DAC_D[0..13]
RN23
RN20
RN29
RN17A 10
RN17B 10
RN17C 10
RN17D 10
RN14A 10
RN14B 10
RN14C 10
RN14D 10
RN28A 10
RN28B 10
RN28C 10
RN28D 10
R61 10
R52 10
PARALLEL CTRL
8x DAC
OFFSET-A
OFFSET-B
DAC8228SRTQ
RST
RSTSEL
CDAC
CLR
OSB/BTC
BUSY
GPIO
DGND
AGND-A
AGND-B
AVSS
PS_DAC_RST
0V Vout after reset
PS_DAC_LDAC
PS_DAC_CLR
straight binary mode
PS_DAC_BUSY_2
pull up at source
RN38A 10
RN38B 10
RN38C 10
RN38D 10
R77
R66
R92 4.7K
R89 4.7K
R86 4.7K
VM_PS_UCTRL_20/25.3C
VM_PS_UCTRL_19/25.3E
VM_PS_UCTRL_18/24.3C
VM_PS_UCTRL_17/24.3E
VM_PS_UCTRL_24/27.3C
VM_PS_UCTRL_23/27.3E
VM_PS_UCTRL_22/26.3C
VM_PS_UCTRL_21/26.3E
2243
33
1143
do not populate
Assembly Variant: standard
Author: Th. Kuehne
Title: ham_v1
Project: Hybrid MMIMO analog module
Document Number:
REV: 1.1
Date: 11.09.18 20:08
Sheet: 13/31

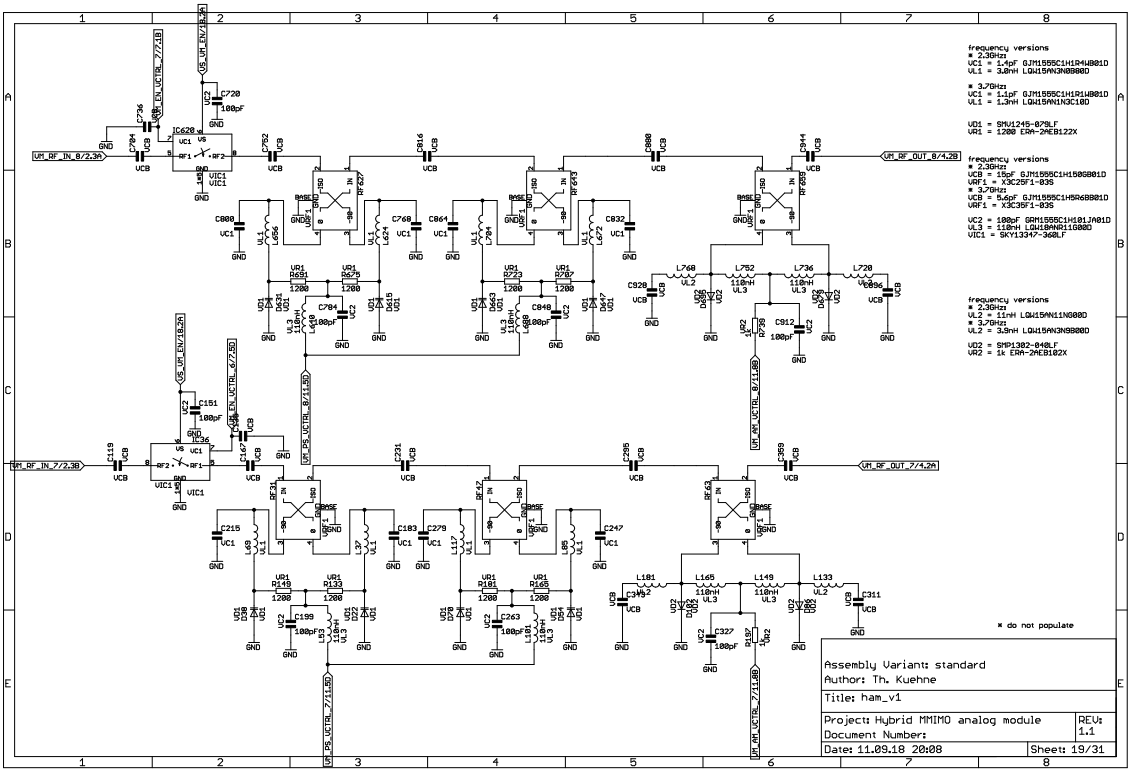

* do not populate
Assembly Variant: standard
Author: Th. Kuehne
Title: ham_v1
Project: Hybrid MMIMO analog module
Document Number:
REV: 1.1
Date: 11.09.18 20:08
Sheet: 14/31

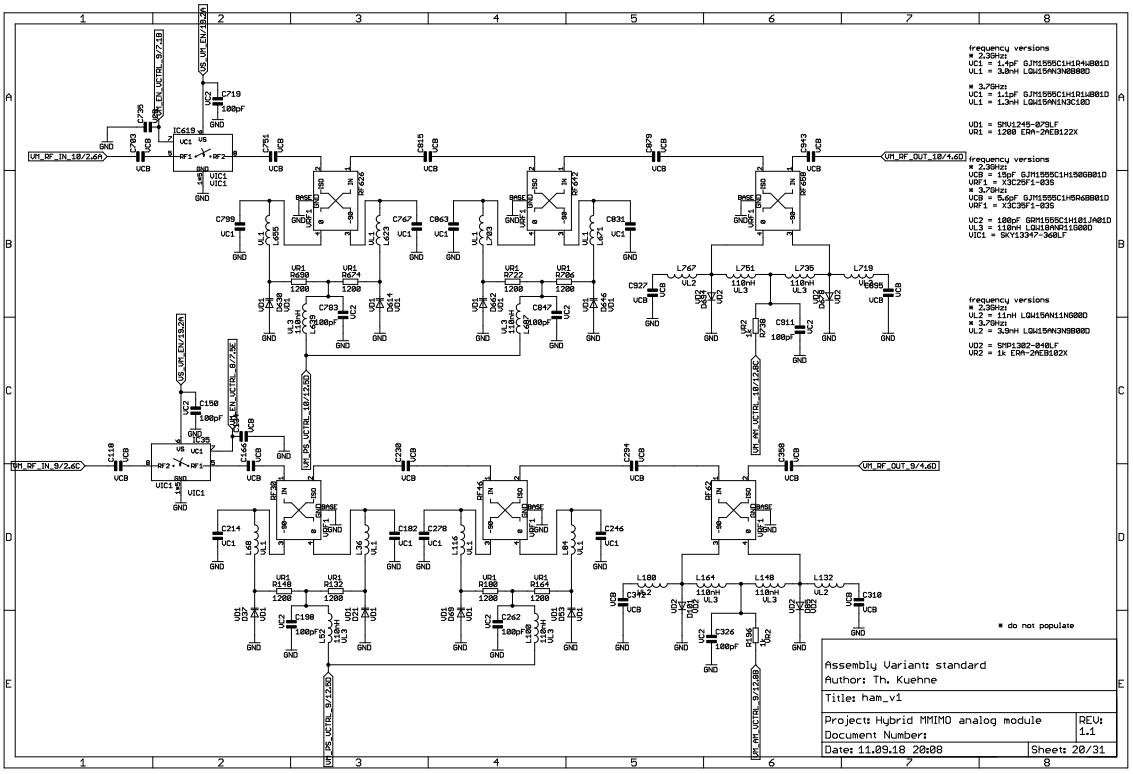

* do not populate
Assembly Variant: standard
Author: Th. Kuehne
Title: ham_v1
Project: Hybrid MMIMO analog module
Document Number:
REV: 1.1
Date: 11.09.18 20:08
Sheet: 15/31

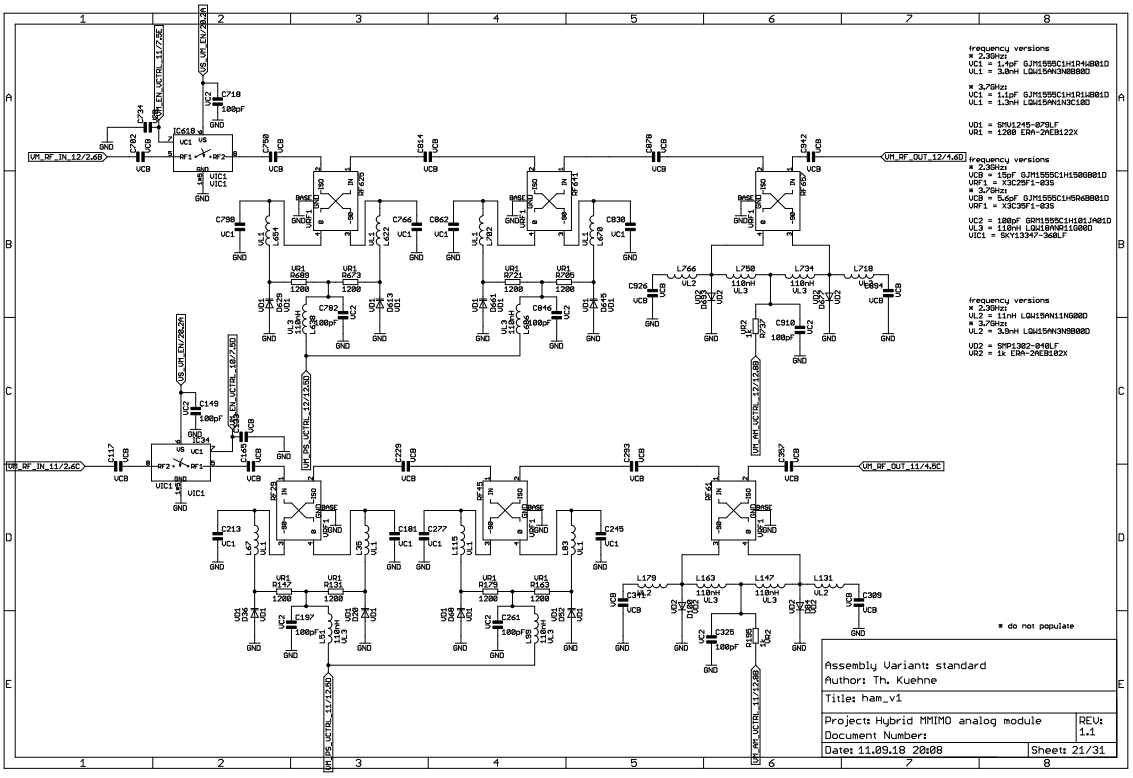

frequency versions
* 2.3GHz:
VC1 = 1.4pF GJM1555C1H1R4WB01D
VL1 = 3.0nH LQW15AN3N0B80D
* 3.7GHz:
VC1 = 1.1pF GJM1555C1H1R1WB01D
VL1 = 1.3nH LQW15AN1N3C10D
VD1 = SMV1245-079LF
VR1 = 120Ω ERA-2AEB122X
frequency versions
* 2.3GHz:
VC3 = 15pF GJM1555C1H150GB01D
VRF1 = X3C25F1-03S
* 3.7GHz:
VC3 = 5.6pF GJM1555C1H5R6BB01D
VRF1 = X3C35F1-03S
VC2 = 100pF GRM1555C1H101JA01D
VL3 = 110nH LQW18AN0R11B80D
VIC1 = SKY13347-360LF
frequency versions
* 2.3GHz:
VL2 = 11nH LQW15AN11N080D
* 3.7GHz:
VL2 = 3.9nH LQW15AN3N9B80D
VD2 = SMP1302-040LF
VR2 = 1k ERA-2AEB102X
* do not populate
Assembly Variant: standard
Author: Th. Kuehne
Title: ham_v1
Project: Hybrid MMIMO analog module
Document Number:
REV: 1.1
Date: 11.09.18 20:08
Sheet: 16/31

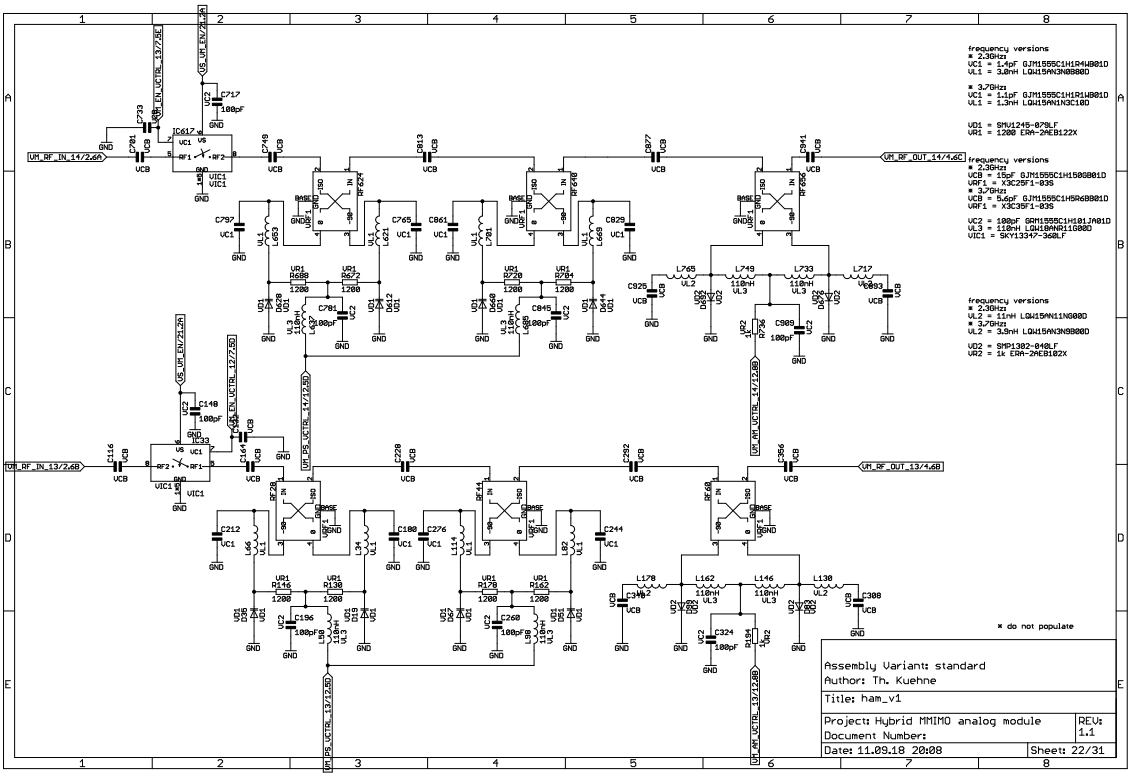

frequency versions
* 2.3GHz:
VC1 = 1.4pF GJM1555C1H1R4WB01D
VL1 = 3.0nH LQW15AN3N0B80D
* 3.7GHz:
VC1 = 1.1pF GJM1555C1H1R1WB01D
VL1 = 1.3nH LQW15AN1N3C10D
VD1 = SMV1245-079LF
VR1 = 120Ω ERA-2AEB122X
frequency versions
* 2.3GHz:
VC3 = 15pF GJM1555C1H150GB01D
VRF1 = X3C25F1-03S
* 3.7GHz:
VC3 = 5.6pF GJM1555C1H5R6BB01D
VRF1 = X3C35F1-03S
VC2 = 100pF GRM1555C1H101JA01D
VL3 = 110nH LQW18AN0R11B80D
VIC1 = SKY13347-360LF
frequency versions
* 2.3GHz:
VL2 = 11nH LQW15AN11N080D
* 3.7GHz:
VL2 = 3.9nH LQW15AN3N9B80D
VD2 = SMP1302-040LF
VR2 = 1k ERA-2AEB102X
* do not populate
Assembly Variant: standard
Author: Th. Kuehne
Title: ham_v1
Project: Hybrid MMIMO analog module
Document Number:
REV: 1.1
Date: 11.09.18 20:08
Sheet: 17/31

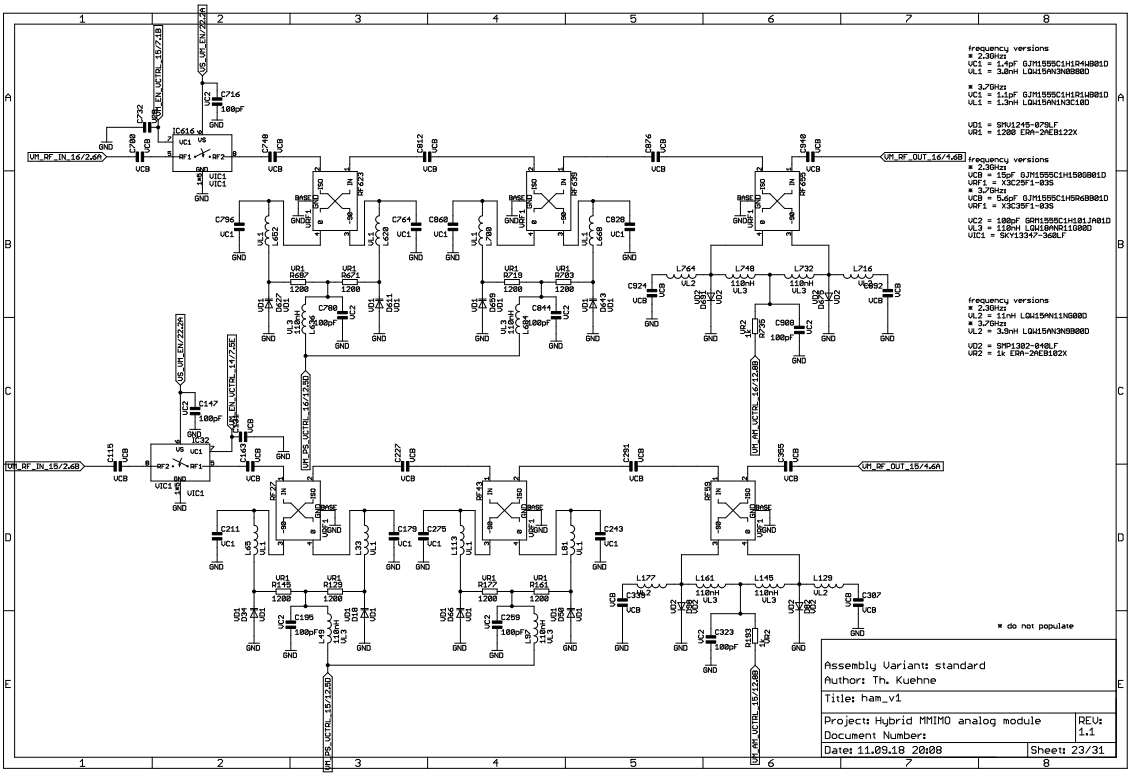

frequency versions
* 2.3GHz
VC1 = 1.4pF GJM1555C1H1R4WB01D
VL1 = 3.0nH LQW15AN3N0B00D
* 3.7GHz
VC1 = 1.1pF GJM1555C1H1R1WB01D
VL1 = 1.3nH LQW15AN1N3C10D
VD1 = SMV1245-079LF
VR1 = 1200 ERA-2AEB122X
frequency versions
* 2.3GHz
VCB = 15pF GJM1555C1H150GB01D
VRF1 = X3C25F1-03S
* 3.7GHz
VCB = 5.6pF GJM1555C1H5R6WB01D
VRF1 = X3C25F1-03S
VC2 = 100pF GRM1555C1H101JA01D
VL3 = 110nH LQW18ANR11J00D
VIC1 = SKY13347-368LF
frequency versions
* 2.3GHz
VL2 = 11nH LQW15AN11NG00D
* 3.7GHz
VL2 = 3.9nH LQW15AN3N9B00D
VD2 = SMP1302-040LF
VR2 = 1k ERA-2AEB102X
* do not populate
Assembly Variant: standard
Author: Th. Kuehne
Title: ham_v1
Project: Hybrid MMIMO analog module
Document Number:
REV: 1.1
Date: 11.09.18 20:08
Sheet: 18/31

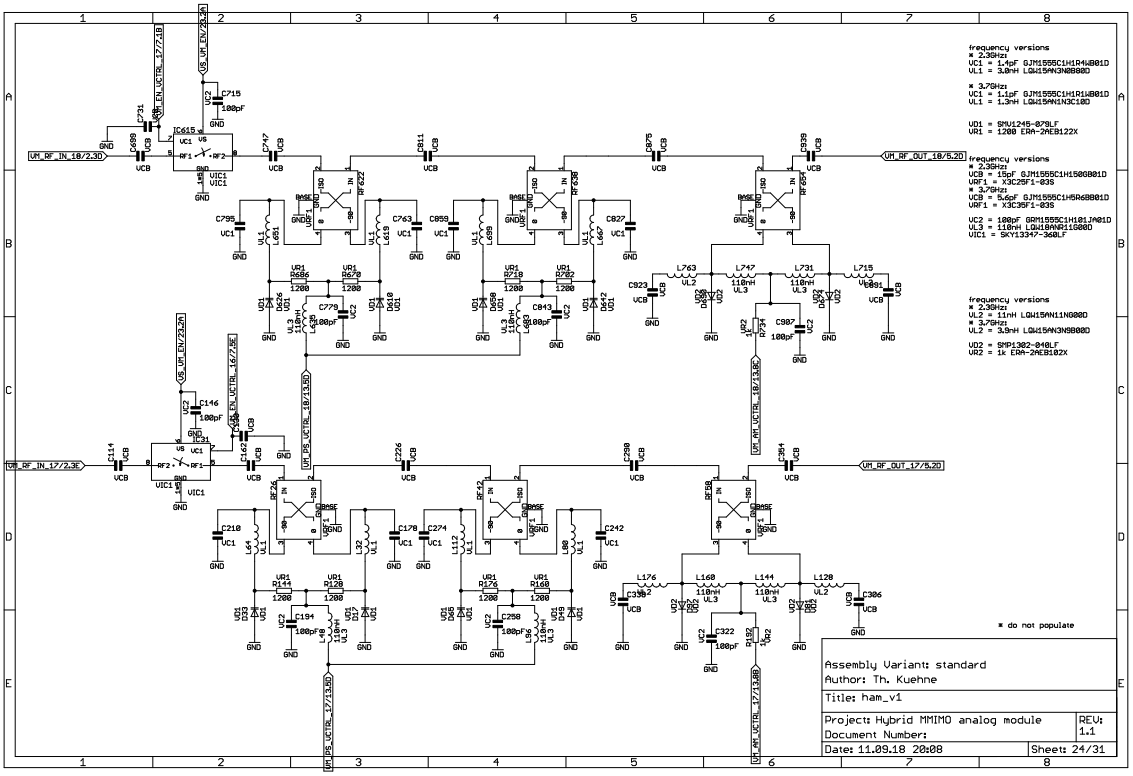

frequency versions
* 2.3GHz
VC1 = 1.4pF GJM1555C1H1R4WB01D
VL1 = 3.0nH LQW15AN3N0B00D
* 3.7GHz
VC1 = 1.1pF GJM1555C1H1R1WB01D
VL1 = 1.3nH LQW15AN1N3C10D
VD1 = SMV1245-079LF
VR1 = 1200 ERA-2AEB122X
frequency versions
* 2.3GHz
VCB = 15pF GJM1555C1H150GB01D
VRF1 = X3C25F1-03S
* 3.7GHz
VCB = 5.6pF GJM1555C1H5R6WB01D
VRF1 = X3C25F1-03S
VC2 = 100pF GRM1555C1H101JA01D
VL3 = 110nH LQW18ANR11J00D
VIC1 = SKY13347-368LF
frequency versions
* 2.3GHz
VL2 = 11nH LQW15AN11NG00D
* 3.7GHz
VL2 = 3.9nH LQW15AN3N9B00D
VD2 = SMP1302-040LF
VR2 = 1k ERA-2AEB102X
* do not populate
Assembly Variant: standard
Author: Th. Kuehne
Title: ham_v1
Project: Hybrid MMIMO analog module
Document Number:
REV: 1.1
Date: 11.09.18 20:08
Sheet: 19/31

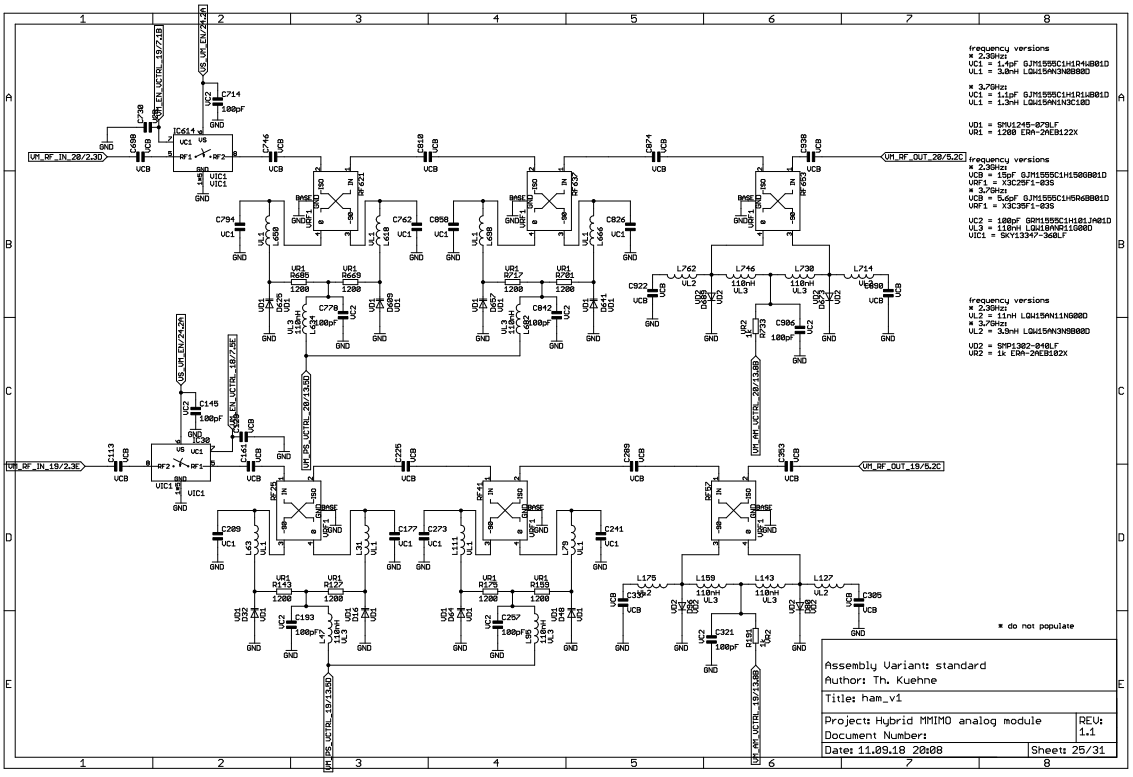

frequency versions
* 2.3GHz:
VC1 = 1.4pF GJM1555C1H1R4WB01D
VL1 = 3.0nH LQW15AN3N0B80D
* 3.7GHz:
VC1 = 1.1pF GJM1555C1H1R1WB01D
VL1 = 1.3nH LQW15AN1N3C10D
VD1 = SMV1245-079LF
VR1 = 1200 ERA-2AEB122X
frequency versions
* 2.3GHz:
VCB = 15pF GJM1555C1H150GB01D
VRF1 = X3C25F1-03S
* 3.7GHz:
VCB = 5.6pF GJM1555C1H5R6BB01D
VRF1 = X3C25F1-03S
VC2 = 100pF GRM1555C1H101JA01D
VL3 = 110nH LQW18AN11G80D
VIC1 = SKY13347-360LF
frequency versions
* 2.3GHz:
VL2 = 11nH LQW15AN11N08D
* 3.7GHz:
VL2 = 3.9nH LQW15AN3N9B80D
VD2 = SMP1302-040LF
VR2 = 1k ERA-2AEB102X
* do not populate
Assembly Variant: standard
Author: Th. Kuehne
Title: ham_v1
Project: Hybrid MMIMO analog module
Document Number:
REV: 1.1
Date: 11.09.18 20:08
Sheet: 20/31

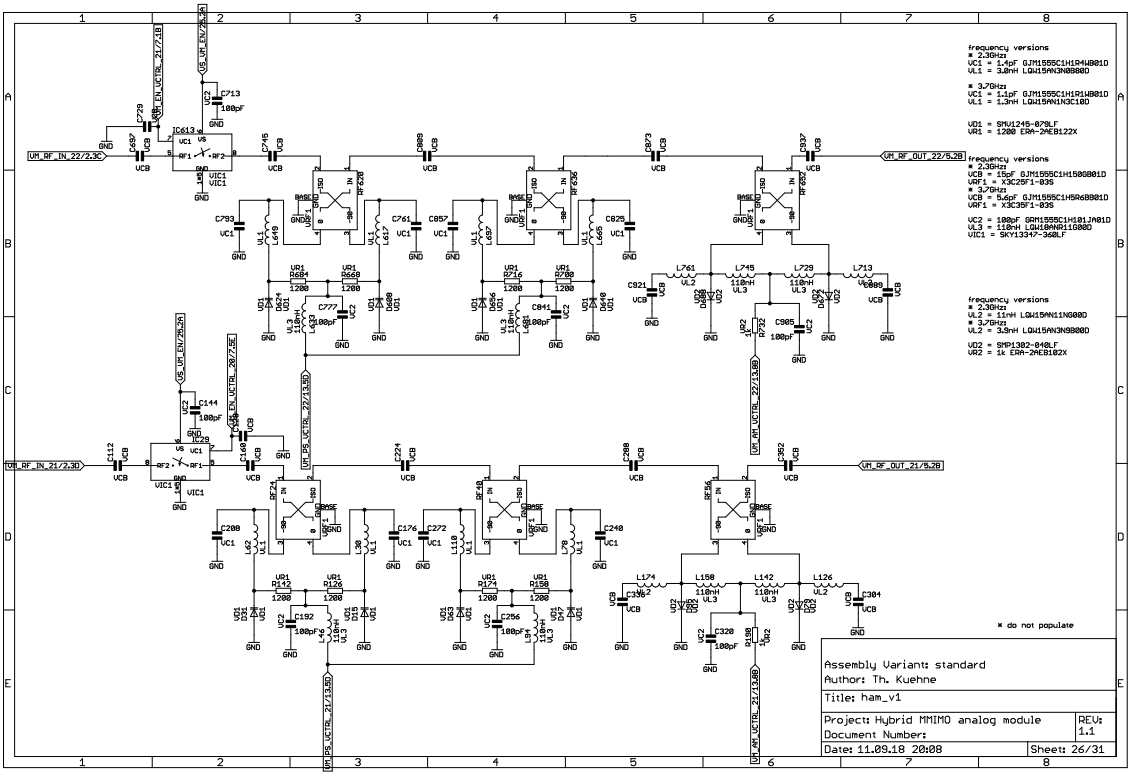

frequency versions
* 2.3GHz:
VC1 = 1.4pF GJM1555C1H1R4WB01D
VL1 = 3.0nH LQW15AN3N0B80D
* 3.7GHz:
VC1 = 1.1pF GJM1555C1H1R1WB01D
VL1 = 1.3nH LQW15AN1N3C10D
VD1 = SMV1245-079LF
VR1 = 1200 ERA-2AEB122X
frequency versions
* 2.3GHz:
VCB = 15pF GJM1555C1H150GB01D
VRF1 = X3C25F1-03S
* 3.7GHz:
VCB = 5.6pF GJM1555C1H5R6BB01D
VRF1 = X3C25F1-03S
VC2 = 100pF GRM1555C1H101JA01D
VL3 = 110nH LQW18AN11G80D
VIC1 = SKY13347-360LF
frequency versions
* 2.3GHz:
VL2 = 11nH LQW15AN11N08D
* 3.7GHz:
VL2 = 3.9nH LQW15AN3N9B80D
VD2 = SMP1302-040LF
VR2 = 1k ERA-2AEB102X
* do not populate
Assembly Variant: standard
Author: Th. Kuehne
Title: ham_v1
Project: Hybrid MMIMO analog module
Document Number:
REV: 1.1
Date: 11.09.18 20:08
Sheet: 21/31

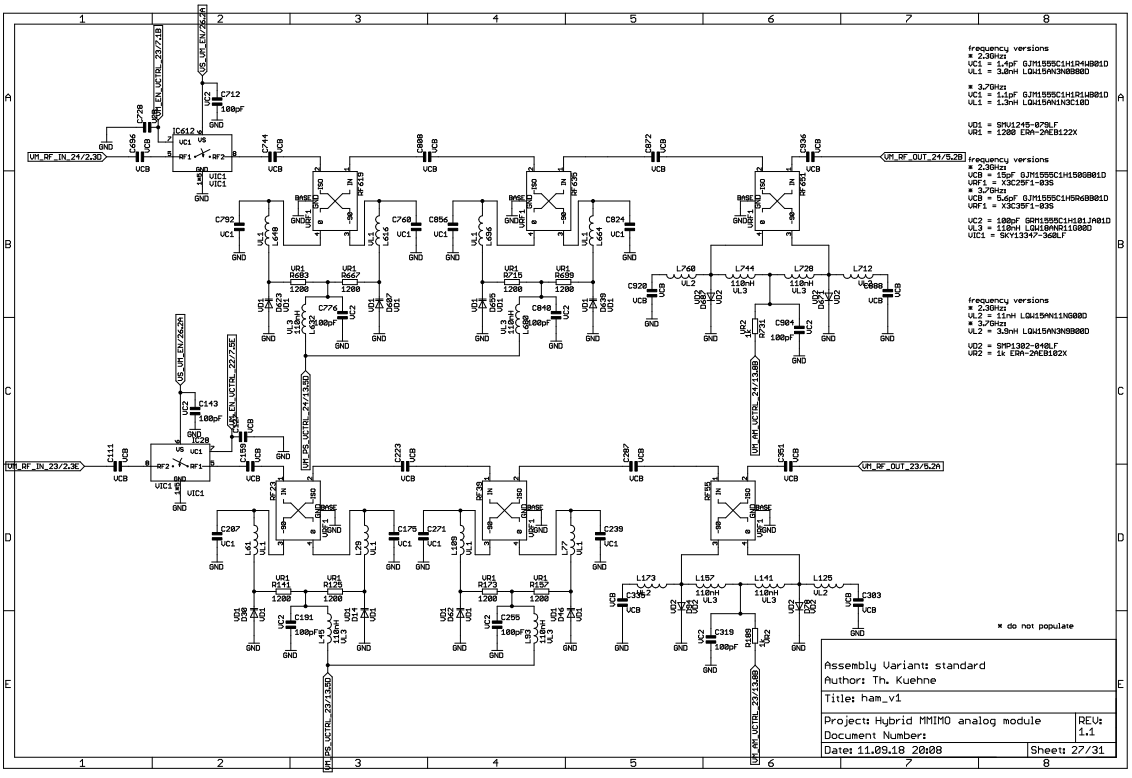

frequency versions
* 2.3GHz
VC1 = 1.4pF GJM1555C1H1R4WB01D
VL1 = 3.0nH LQW15AN3N0B80
* 3.7GHz
VC1 = 1.1pF GJM1555C1H1R1WB01D
VL1 = 1.3nH LQW15AN1N3C10
VD1 = SMV1245-079LF
VR1 = 120Ω ERA-2AEB122X
frequency versions
* 2.3GHz
VCB = 15pF GJM1555C1H150B01D
VR1 = X3C25F1-03S
* 3.7GHz
VCB = 5.6pF GJM1555C1H5R6B001D
VR1 = X3C35F1-03S
VC2 = 100pF GRM1555C1H101JA01D
VL3 = 110nH LQW18ANR11G00D
VIC1 = SKY13347-368LF
frequency versions
* 2.3GHz
VL2 = 11nH LQW15AN11NG00D
* 3.7GHz
VL2 = 3.3nH LQW15AN3N9B00D
VD2 = SMP1302-040LF
VR2 = 1k ERA-2AEB102X
* do not populate
Assembly Variant: standard
Author: Th. Kuehne
Title: ham_v1
Project: Hybrid MMIMO analog module
Document Number:
REV: 1.1
Date: 11.09.18 20:08
Sheet: 22/31

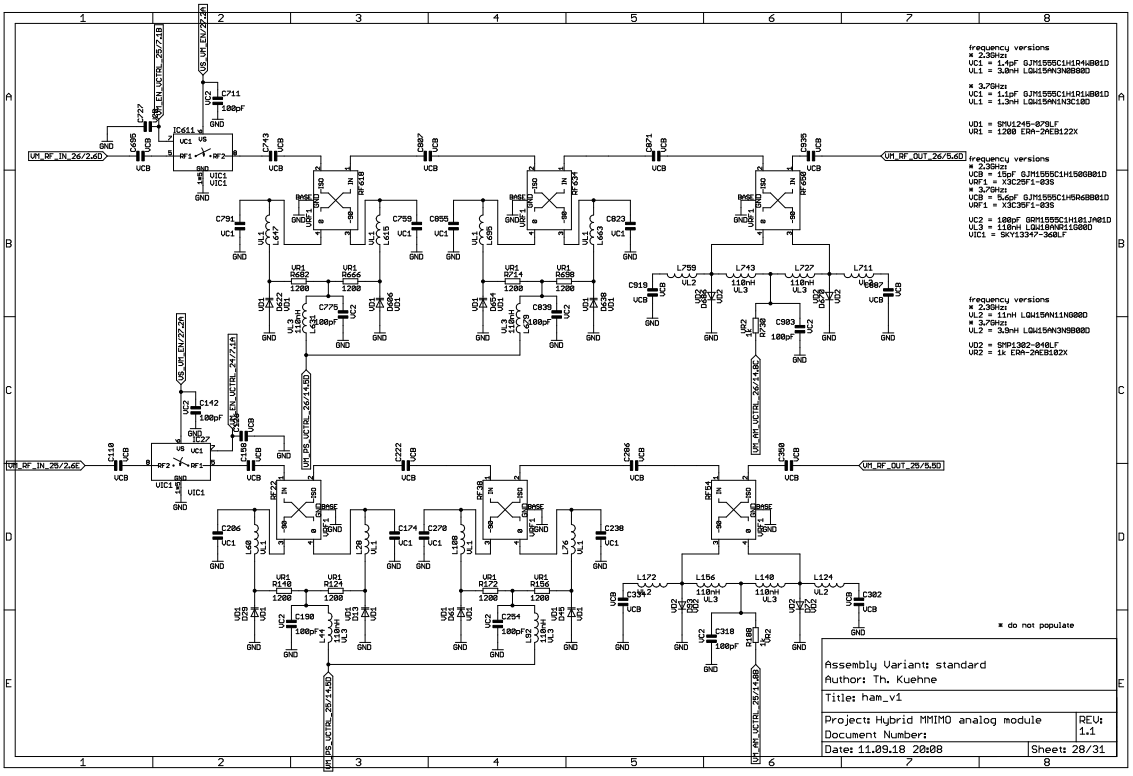

frequency versions
* 2.3GHz
VC1 = 1.4pF GJM1555C1H1R4WB01D
VL1 = 3.0nH LQW15AN3N0B80
* 3.7GHz
VC1 = 1.1pF GJM1555C1H1R1WB01D
VL1 = 1.3nH LQW15AN1N3C10
VD1 = SMV1245-079LF
VR1 = 120Ω ERA-2AEB122X
frequency versions
* 2.3GHz
VCB = 15pF GJM1555C1H150B01D
VR1 = X3C25F1-03S
* 3.7GHz
VCB = 5.6pF GJM1555C1H5R6B001D
VR1 = X3C35F1-03S
VC2 = 100pF GRM1555C1H101JA01D
VL3 = 110nH LQW18ANR11G00D
VIC1 = SKY13347-368LF
frequency versions
* 2.3GHz
VL2 = 11nH LQW15AN11NG00D
* 3.7GHz
VL2 = 3.3nH LQW15AN3N9B00D
VD2 = SMP1302-040LF
VR2 = 1k ERA-2AEB102X
* do not populate
Assembly Variant: standard
Author: Th. Kuehne
Title: ham_v1
Project: Hybrid MMIMO analog module
Document Number:
REV: 1.1
Date: 11.09.18 20:08
Sheet: 23/31

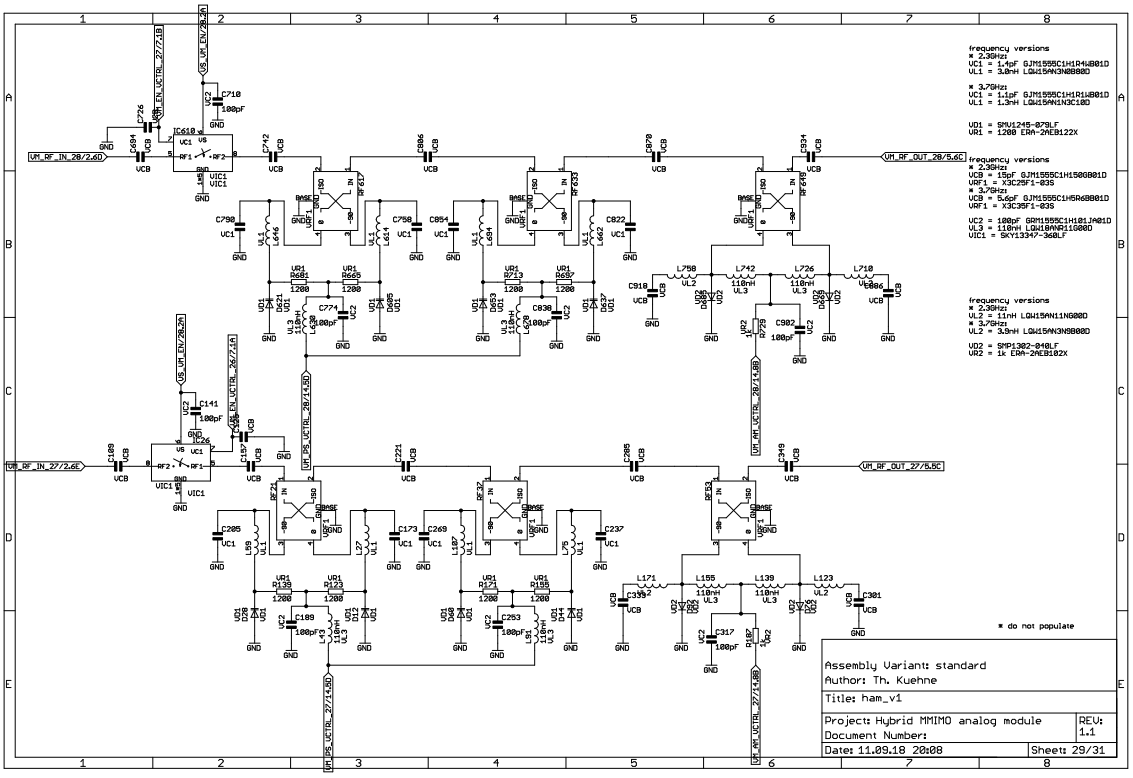

frequency versions
= 2.3GHz
VC1 = 1.4pF GJM1555C1H1R4WB01D
VL1 = 3.0nH LQW15AN3N0B80D
= 3.7GHz
VC1 = 1.1pF GJM1555C1H1R1WB01D
VL1 = 1.3nH LQW15AN1N3C10D
VD1 = SMV1245-079LF
VR1 = 120Ω ERA-2AEB122X
frequency versions
= 2.3GHz
VCB = 15pF GJM1555C1H150GB01D
VRF1 = X3C25F1-03S
= 3.7GHz
VCB = 5.6pF GJM1555C1H5R6BB01D
VRF1 = X3C35F1-03S
VC2 = 100pF GRM1555C1H101JA01D
VL3 = 110nH LQW18AN11G80D
VIC1 = SKY13347-360LF
frequency versions
= 2.3GHz
VL2 = 11nH LQW15AN11N080D
= 3.7GHz
VL2 = 3.9nH LQW15AN3N9B80D
VD2 = SMP1302-040LF
VR2 = 1k ERA-2AEB102X
* do not populate
Assembly Variant: standard
Author: Th. Kuehne
Title: ham_v1
Project: Hybrid MMIMO analog module
Document Number:
REV: 1.1
Date: 11.09.18 20:08
Sheet: 24/31

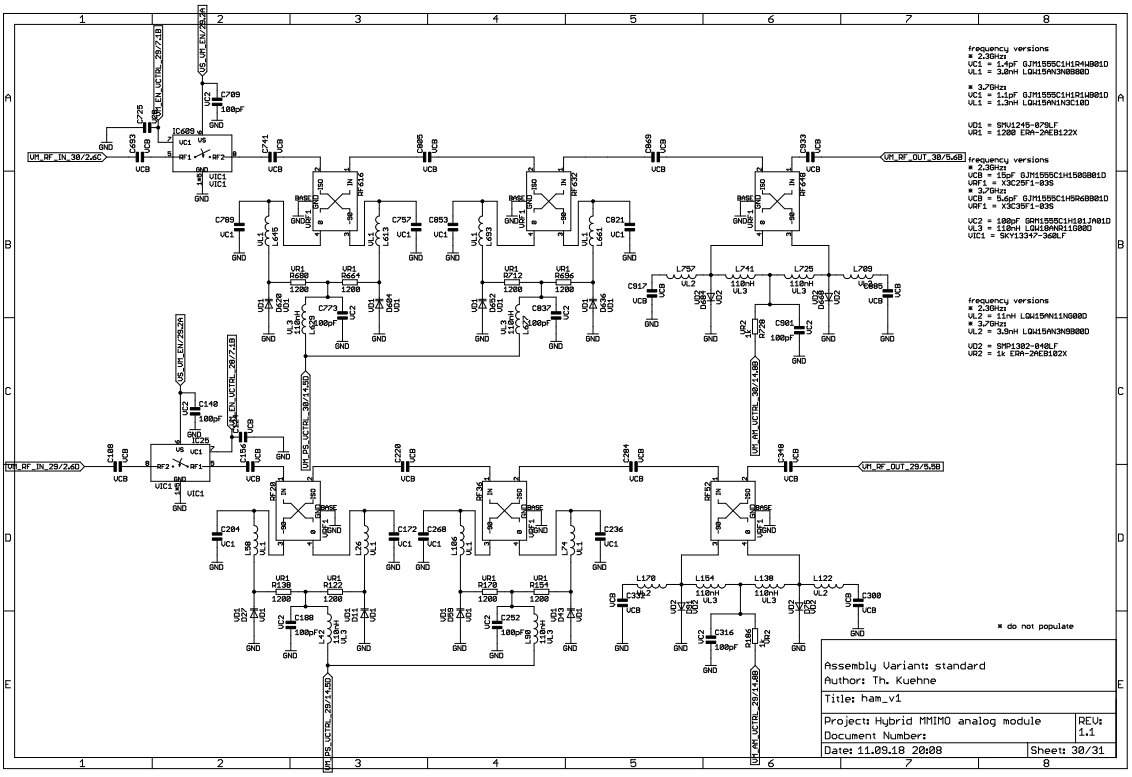

frequency versions
= 2.3GHz
VC1 = 1.4pF GJM1555C1H1R4WB01D
VL1 = 3.0nH LQW15AN3N0B80D
= 3.7GHz
VC1 = 1.1pF GJM1555C1H1R1WB01D
VL1 = 1.3nH LQW15AN1N3C10D
VD1 = SMV1245-079LF
VR1 = 120Ω ERA-2AEB122X
frequency versions
= 2.3GHz
VCB = 15pF GJM1555C1H150GB01D
VRF1 = X3C25F1-03S
= 3.7GHz
VCB = 5.6pF GJM1555C1H5R6BB01D
VRF1 = X3C35F1-03S
VC2 = 100pF GRM1555C1H101JA01D
VL3 = 110nH LQW18AN11G80D
VIC1 = SKY13347-360LF
frequency versions
= 2.3GHz
VL2 = 11nH LQW15AN11N080D
= 3.7GHz
VL2 = 3.9nH LQW15AN3N9B80D
VD2 = SMP1302-040LF
VR2 = 1k ERA-2AEB102X
* do not populate
Assembly Variant: standard
Author: Th. Kuehne
Title: ham_v1
Project: Hybrid MMIMO analog module
Document Number:
REV: 1.1
Date: 11.09.18 20:08
Sheet: 25/31

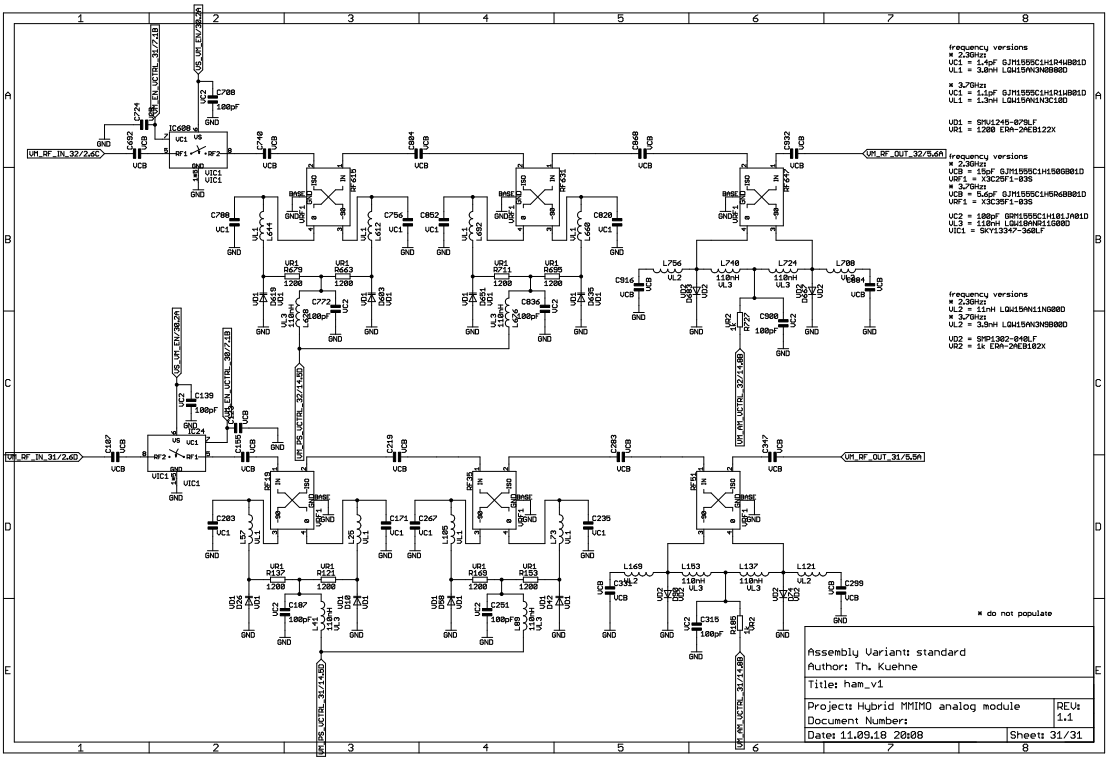

frequency versions
* 2.3GHz
VC1 = 1.4pF GJM1555C1H1R4WB01D
VL1 = 3.0nH LQW15AN3N0B80
* 3.7GHz
VC1 = 1.1pF GJM1555C1H1R1WB01D
VL1 = 1.3nH LQW15AN1N3C10
VD1 = SMV1245-079LF
VR1 = 1200 ERA-2AEB122X
VH_RF_IN_22/2.3C
VH_RF_OUT_22/5.2B
frequency versions
* 2.3GHz
VCB = 15pF GJM1555C1H150B01D
VR1 = X3C25F1-03S
* 3.7GHz
VCB = 5.6pF GJM1555C1H5R6B01D
VR1 = X3C35F1-03S
VC2 = 100pF GRM1555C1H101JA01D
VL3 = 110nH LQW18ANR11G00D
VIC1 = SKY13347-368LF
frequency versions
* 2.3GHz
VL2 = 11nH LQW15AN11NG00D
* 3.7GHz
VL2 = 3.9nH LQW15AN3N9B00D
VD2 = SMP1302-040LF
VR2 = 1k ERA-2AEB102X
VH_RF_IN_21/2.3D
VH_RF_OUT_21/5.2B
* do not populate
Assembly Variant: standard
Author: Th. Kuehne
Title: ham_v1
Project: Hybrid MMIMO analog module
Document Number:
REV: 1.1
Date: 11.09.18 20:08
Sheet: 26/31


frequency versions
* 2.3GHz
VC1 = 1.4pF GJM1555C1H1R4WB01D
VL1 = 3.0nH LQW15AN3N0B80
* 3.7GHz
VC1 = 1.1pF GJM1555C1H1R1WB01D
VL1 = 1.3nH LQW15AN1N3C10
VD1 = SMV1245-079LF
VR1 = 1200 ERA-2AEB122X
VH_RF_IN_24/2.3D
VH_RF_OUT_24/5.2B
frequency versions
* 2.3GHz
VCB = 15pF GJM1555C1H150B01D
VR1 = X3C25F1-03S
* 3.7GHz
VCB = 5.6pF GJM1555C1H5R6B01D
VR1 = X3C35F1-03S
VC2 = 100pF GRM1555C1H101JA01D
VL3 = 110nH LQW18ANR11G00D
VIC1 = SKY13347-368LF
frequency versions
* 2.3GHz
VL2 = 11nH LQW15AN11NG00D
* 3.7GHz
VL2 = 3.9nH LQW15AN3N9B00D
VD2 = SMP1302-040LF
VR2 = 1k ERA-2AEB102X
VH_RF_IN_23/2.3E
VH_RF_OUT_23/5.2A
* do not populate
Assembly Variant: standard
Author: Th. Kuehne
Title: ham_v1
Project: Hybrid MMIMO analog module
Document Number:
REV: 1.1
Date: 11.09.18 20:08
Sheet: 27/31


frequency versions
= 2.3GHz:
VC1 = 1.4pF GJM1555C1H1R4WB01D
VL1 = 3.0nH LQW15AN3N0B80D
= 3.7GHz:
VC1 = 1.1pF GJM1555C1H1R1WB01D
VL1 = 1.3nH LQW15AN1N3C10D
VD1 = SMV1245-079LF
VR1 = 120Ω ERA-2AEB122X
frequency versions
= 2.3GHz:
VCB = 15pF GJM1555C1H150GB01D
VRF1 = X3C25F1-03S
= 3.7GHz:
VCB = 5.6pF GJM1555C1H5R6BB01D
VRF1 = X3C35F1-03S
VC2 = 100pF GRM1555C1H101JA01D
VL3 = 110nH LQW18AN11G80D
VIC1 = SKY13347-360LF
frequency versions
= 2.3GHz:
VL2 = 11nH LQW15AN11N080D
= 3.7GHz:
VL2 = 3.9nH LQW15AN3N9B80D
VD2 = SMP1302-040LF
VR2 = 1k ERA-2AEB102X
* do not populate
Assembly Variant: standard
Author: Th. Kuehne
Title: ham_v1
Project: Hybrid MMIMO analog module
Document Number:
REV: 1.1
Date: 11.09.18 20:08
Sheet: 28/31


frequency versions
= 2.3GHz:
VC1 = 1.4pF GJM1555C1H1R4WB01D
VL1 = 3.0nH LQW15AN3N0B80D
= 3.7GHz:
VC1 = 1.1pF GJM1555C1H1R1WB01D
VL1 = 1.3nH LQW15AN1N3C10D
VD1 = SMV1245-079LF
VR1 = 120Ω ERA-2AEB122X
frequency versions
= 2.3GHz:
VCB = 15pF GJM1555C1H150GB01D
VRF1 = X3C25F1-03S
= 3.7GHz:
VCB = 5.6pF GJM1555C1H5R6BB01D
VRF1 = X3C35F1-03S
VC2 = 100pF GRM1555C1H101JA01D
VL3 = 110nH LQW18AN11G80D
VIC1 = SKY13347-360LF
frequency versions
= 2.3GHz:
VL2 = 11nH LQW15AN11N080D
= 3.7GHz:
VL2 = 3.9nH LQW15AN3N9B80D
VD2 = SMP1302-040LF
VR2 = 1k ERA-2AEB102X
* do not populate
Assembly Variant: standard
Author: Th. Kuehne
Title: ham_v1
Project: Hybrid MMIMO analog module
Document Number:
REV: 1.1
Date: 11.09.18 20:08
Sheet: 29/31


frequency versions
* 2.3GHz
VC1 = 1.4pF GJM1555C1H1R4WB01D
VL1 = 3.0nH LQW15AN3N0B80D
* 3.7GHz
VC1 = 1.1pF GJM1555C1H1R1WB01D
VL1 = 1.3nH LQW15AN1N3C10D
VD1 = SMV1245-079LF
VR1 = 120Ω ERA-2AEB122X
frequency versions
* 2.3GHz
VCB = 15pF GJM1555C1H150GB01D
VRF1 = X3C25F1-03S
* 3.7GHz
VCB = 5.6pF GJM1555C1H5R6BB01D
VRF1 = X3C35F1-03S
VC2 = 100pF GRM1555C1H101JA01D
VL3 = 110nH LQW18ANR11G00D
VIC1 = SKY13347-360LF
frequency versions
* 2.3GHz
VL2 = 11nH LQW15AN11NG00D
* 3.7GHz
VL2 = 3.9nH LQW15AN3N9B00D
VD2 = SMP1302-040LF
VR2 = 1k ERA-2AEB102X
* do not populate
Assembly Variant: standard
Author: Th. Kuehne
Title: ham_v1
Project: Hybrid MMIMO analog module
Document Number:
REV: 1.1
Date: 11.09.18 20:08
Sheet: 30/31
VM_RF_IN_30/2.6G
VM_RF_OUT_30/5.6G
VM_RF_IN_29/2.6G
VM_RF_OUT_29/5.5G
VM_EN_VCTRL_29/7.1G
VM_PS_VCTRL_29/14.5G
VM_AM_VCTRL_29/14.5G


frequency versions
* 2.3GHz
VC1 = 1.4pF GJM1555C1H1R4WB01D
VL1 = 3.0nH LQW15AN3N0B80D
* 3.7GHz
VC1 = 1.1pF GJM1555C1H1R1WB01D
VL1 = 1.3nH LQW15AN1N3C10D
VD1 = SMV1245-079LF
VR1 = 120Ω ERA-2AEB122X
frequency versions
* 2.3GHz
VCB = 15pF GJM1555C1H150GB01D
VRF1 = X3C25F1-03S
* 3.7GHz
VCB = 5.6pF GJM1555C1H5R6BB01D
VRF1 = X3C35F1-03S
VC2 = 100pF GRM1555C1H101JA01D
VL3 = 110nH LQW18ANR11G00D
VIC1 = SKY13347-360LF
frequency versions
* 2.3GHz
VL2 = 11nH LQW15AN11NG00D
* 3.7GHz
VL2 = 3.9nH LQW15AN3N9B00D
VD2 = SMP1302-040LF
VR2 = 1k ERA-2AEB102X
* do not populate
Assembly Variant: standard
Author: Th. Kuehne
Title: ham_v1
Project: Hybrid MMIMO analog module
Document Number:
REV: 1.1
Date: 11.09.18 20:08
Sheet: 31/31
VM_RF_IN_32/2.6G
VM_RF_OUT_32/5.6A
VM_RF_IN_31/2.6G
VM_RF_OUT_31/5.5A
VM_EN_VCTRL_32/7.1G
VM_PS_VCTRL_32/14.5G
VM_AM_VCTRL_32/14.5G

# Bibliography


[1] P. Cerwall, P. Jonsson, S. Carson, *et al.*, "Ericsson Mobility Report", Ericsson, White Paper, Jun. 2021. [Online]. Available: https://www.ericsson.com/4a03c2/assets/local/mobility-report/documents/2021/june-2021-ericsson-mobility-report.pdf.

[2] "Cisco Annual Internet Report (2018—2023)", Cisco, White Paper, Mar. 2020. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.pdf.

[3] D. Gesbert, M. Kountouris, R. W. Heath, C.-b. Chae, and T. Salzer, "Shifting the MIMO Paradigm", *IEEE Signal Processing Magazine*, vol. 24, no. 5, pp. 36–46, 2007. DOI: 10.1109/MSP.2007.904815.

[4] R. W. Heath Jr. and A. Lozano, *Foundations of MIMO Communication*. Cambridge University Press, Dec. 2018. DOI: 10.1017/9781139049276.

[5] R. Karmakar, S. Chattopadhyay, and S. Chakraborty, "Impact of IEEE 802.11n/ac PHY/MAC High Throughput Enhancements on Transport and Application Protocols—A Survey", *IEEE Communications Surveys and Tutorials*, vol. 19, no. 4, pp. 2050–2091, 2017. DOI: 10.1109/comst.2017.2745052.

[6] O. Bejarano, E. Knightly, and M. Park, "IEEE 802.11ac: from channelization to multi-user MIMO", *IEEE Communications Magazine*, vol. 51, no. 10, pp. 84–90, Oct. 2013, ISSN: 0163-6804. DOI: 10.1109/mcom.2013.6619570.

[7] E. Khorov, A. Kiryanov, A. Lyakhov, and G. Bianchi, "A Tutorial on IEEE 802.11ax High Efficiency WLANs", *IEEE Communications Surveys and Tutorials*, vol. 21, no. 1, pp. 197–216, 2019, ISSN: 1553-877X. DOI: 10.1109/comst.2018.2871099.

[8] Qualcomm, "802.11ac MU-MIMO: Bridging the MIMO Gap in Wi-Fi", Qualcomm Atheros, Inc., White Paper, Jan. 2015. [Online]. Available: https://www.qualcomm.com/media/documents/files/802-11ac-mu-mimo-bridging-the-mimo-gap-in-wi-fi.pdf.

[9] C. Lim, T. Yoo, B. Clerckx, B. Lee, and B. Shim, "Recent trend of multiuser MIMO in LTE-advanced", *IEEE Communications Magazine*, vol. 51, no. 3, pp. 127–135, 2013. DOI: 10.1109/MCOM.2013.6476877.

[10] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G", *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, 2014. DOI: 10.1109/MCOM.2014.6736746.

[11] T. L. Marzetta, "Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas", *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010, ISSN: 1536-1276. DOI: 10.1109/TWC.2010.092810.091092.

[12] Nokia, *Nokia 4.9G Massive MIMO enables 5G-like user experiences.* Executive Summary, 2018.

[13] W. B. Hasan, P. Harris, A. Doufexi, and M. Beach, "Real-Time Maximum Spectral Efficiency for Massive MIMO and its Limits", *IEEE Access*, vol. 6, pp. 46 122–46 133, 2018, ISSN: 2169-3536. DOI: 10.1109/access.2018.2866094.

[14] A. F. Molisch, V. V. Ratnam, S. Han, *et al.*, "Hybrid Beamforming for Massive MIMO: A Survey", *IEEE Communications Magazine*, vol. 55, no. 9, pp. 134–141, 2017, ISSN: 0163-6804. DOI: 10.1109/mcom.2017.1600400.

[15] T. S. Rappaport, Y. Xing, G. R. MacCartney, A. F. Molisch, E. Mellios, and J. Zhang, "Overview of Millimeter Wave Communications for Fifth-Generation (5G) Wireless Networks —With a Focus on Propagation Models", *IEEE Transactions on Antennas and Propagation*, vol. 65, no. 12, pp. 6213–6230, Dec. 2017. DOI: 10.1109/TAP.2017.2734243.

[16] 3GPP TR 38.101-1, "NR; User Equipment (UE) radio transmission and reception; Part 1: Range 1 Standalone", Tech. Rep. v17.2.0, 2021. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3283.

[17] R. W. Heath Jr., N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An Overview of Signal Processing Techniques for Millimeter Wave MIMO Systems", *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 436–453, Apr. 2016, ISSN: 1941-0484. DOI: 10.1109/JSTSP.2016.2523924.

[18] A. Li and C. Masouros, "Hybrid Analog-Digital Millimeter-Wave MU-MIMO Transmission with Virtual Path Selection", *IEEE Communications Letters*, vol. 21, no. 2, pp. 438–441, 2017, ISSN: 1089-7798. DOI: 10.1109/LCOMM.2016.2621741.

[19] V. Raghavan, S. Subramanian, J. Cezanne, A. Sampath, O. H. Koymen, and J. Li, "Single-User Versus Multi-User Precoding for Millimeter Wave MIMO Systems", *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 6, pp. 1387–1401, Jun. 2017, ISSN: 1558-0008. DOI: 10.1109/JSAC.2017.2687798.

[20] T. Kuehne, X. Song, G. Caire, *et al.*, "Performance Simulation of a 5G Hybrid Beamforming Millimeter-Wave System", in *24th International ITG Workshop on Smart Antennas (WSA 2020)*, Hamburg, Germany, Feb. 2020.

[21] X. Song, "Millimeter wave wireless communication: initial acquisition, data communication and relay network investigation", Ph.D. dissertation, Technische Universität Berlin, Berlin, 2020. DOI: 10.14279/depositonce-10625.

[22] T. E. Bogale, L. B. Le, A. Haghighat, and L. Vandendorpe, "On the Number of RF Chains and Phase Shifters, and Scheduling Design With Hybrid Analog-Digital Beamforming", *IEEE Transactions on Wireless Communications*, vol. 15, no. 5, pp. 3311–3326, May 2016. DOI: 10.1109/TWC.2016.2519883.

[23] R. Méndez-Rial, C. Rusu, N. González-Prelcic, A. Alkhateeb, and R. W. Heath, "Hybrid MIMO Architectures for Millimeter Wave Communications: Phase Shifters or Switches?", *IEEE Access*, vol. 4, pp. 247–267, 2016, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2015.2514261.

[24] 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*, vol. 17, no. 8, pp. 5465–5479, 2018. DOI: 10.1109/TWC.2018.2844207.

[25] X. Song, T. Kühne, and G. Caire, "Fully-Connected vs. Sub-Connected Hybrid Precoding Architectures for mmWave MU-MIMO", in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, May 2019. DOI: 10.1109/ICC.2019.8761521.

[26] N. N. Moghadam, G. Fodor, M. Bengtsson, and D. J. Love, "On the Energy Efficiency of MIMO Hybrid Beamforming for Millimeter-Wave Systems With Nonlinear Power Amplifiers", *IEEE Transactions on Wireless Communications*, vol. 17, no. 11, pp. 7208–7221, 2018. DOI: 10.1109/TWC.2018.2865786.

[27] X. Song, T. Kühne, and G. Caire, "Fully-/Partially-Connected Hybrid Beamforming Architectures for mmWave MU-MIMO", *IEEE Transactions on Wireless Communications*, vol. 19, no. 3, pp. 1754–1769, Mar. 2020, ISSN: 1536-1276. DOI: 10.1109/twc.2019.2957227.

[28] A. Alkhateeb, G. Leus, and R. W. Heath Jr., "Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?", in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2015, pp. 2909–2913. DOI: 10.1109/ICASSP.2015.7178503.

[29] 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*, vol. 35, no. 9, pp. 1996–2009, 2017. DOI: 10.1109/JSAC.2017.2720856.

[30] S. Haghighatshoar and G. Caire, "Massive MIMO Channel Subspace Estimation From Low-Dimensional Projections", *IEEE Transactions on Signal Processing*, vol. 65, no. 2, pp. 303–318, Jan. 2017, ISSN: 1053-587X. DOI: 10.1109/TSP.2016.2616336.

[31] X. Song, S. Haghighatshoar, and G. Caire, "A Scalable and Statistically Robust Beam Alignment Technique for Millimeter-Wave Systems", *IEEE Transactions on Wireless Communications*, vol. 17, no. 7, pp. 4792–4805, Jul. 2018, ISSN: 1536-1276. DOI: 10.1109/twc.2018.2831697.

[32] L. Liang, W. Xu, and X. Dong, "Low-Complexity Hybrid Precoding in Massive Multiuser MIMO Systems", *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 653–656, Dec. 2014, ISSN: 2162-2337. DOI: 10.1109/LWC.2014.2363831.

[33] S. S. Ioushua and Y. C. Eldar, "A Family of Hybrid Analog-Digital Beamforming Methods for Massive MIMO Systems", *IEEE Transactions on Signal Processing*, vol. 67, no. 12, pp. 3243–3257, 2019. DOI: 10.1109/TSP.2019.2911255.

[34] R. Gomes, L. Sismeiro, C. Ribeiro, *et al.*, "Will COTS RF Front-Ends Really Cope With 5G Requirements at mmWave?", *IEEE Access*, vol. 6, pp. 38 745–38 769, 2018. DOI: 10.1109/ACCESS.2018.2851781.

[35] T. Kühne and G. Caire, "An Analog Module for Hybrid Massive MIMO Testbeds Demonstrating Beam Alignment Algorithms", in *22nd International ITG Workshop on Smart Antennas (WSA 2018)*, Bochum, Germany, Mar. 2018.

[36] T. Kühne, G. Caire, and X. Song, "Signal processing algorithms and specifications", SERENA, research rep., Jan. 2018. DOI: 10.5281/zenodo.3240455.

[37] K. Andersson and T. Kühne, "Proof of Concept Platform and Front end Specifications", SERENA, research rep., Jun. 2019. DOI: 10.5281/zenodo.3240304.

[38] T. Kühne, P. Gawłowicz, A. Zubow, F. Dressler, and G. Caire, "Demo: Bringing Hybrid Analog-Digital Beamforming to Commercial MU-MIMO WiFi Networks", in *26th ACM International Conference on Mobile Computing and Networking (MobiCom 2020)*, London, United Kingdom, Sep. 2020, ISBN: 978-1-4503-7085-1. DOI: 10.1145/3372224.3417320.

[39] T. Kühne, P. Gawłowicz, A. Zubow, F. Dressler, and G. Caire, "Hybrid Analog-Digital Beamforming: Unlocking the Real MU-MIMO Potential of Commodity WiFi", *(to be submitted)*, 2021.

[40] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel", *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1691–1706, Jul. 2003, ISSN: 0018-9448. DOI: 10.1109/TIT.2003.813523.

[41] H. Huh, A. M. Tulino, and G. Caire, "Network MIMO With Linear Zero-Forcing Beamforming: Large System Analysis, Impact of Channel Estimation, and Reduced-Complexity Scheduling", *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 2911–2934, May 2012, ISSN: 0018-9448. DOI: 10.1109/TIT.2011.2178230.

[42] M. R. A. Khandaker and K.-K. Wong, "Chapter 8 - Signal processing for massive MIMO communications", in *Academic Press Library in Signal Processing, Volume 7*, R. Chellappa and S. Theodoridis, Eds., Academic Press, 2018, pp. 367–401, ISBN: 978-0-12-811887-0. DOI: 10.1016/B978-0-12-811887-0.00008-0.

[43] T. L. Marzetta, E. G. Larsson, and H. Yang, *Fundamentals of Massive MIMO*. Cambridge University Press, Nov. 2016, 240 pp., ISBN: 9781107175570. [Online]. Available: https://www.ebook.de/de/product/26457421/thomas_l_marzetta_erik_g_larsson_hong_yang_fundamentals_of_massive_mimo.html.

[44] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A Vector-Perturbation Technique for Near-Capacity Multiantenna Multiuser Communication—Part I: Channel Inversion and Regularization", *IEEE Transactions on Communications*, vol. 53, no. 1, pp. 195–202, Jan. 2005, ISSN: 0090-6778. DOI: 10.1109/tcomm.2004.840638.

[45] A. Adhikary, J. Nam, J. Y. Ahn, and G. Caire, "Joint Spatial Division and Multiplexing — The Large-Scale Array Regime", *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013, ISSN: 0018-9448. DOI: 10.1109/TIT.2013.2269476.

[46] T. L. Marzetta, "How Much Training is Required for Multiuser Mimo?", in *2006 Fortieth Asilomar Conference on Signals, Systems and Computers*, Oct. 2006, pp. 359–363. DOI: 10.1109/ACSSC.2006.354768.

[47] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems", *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, Feb. 2014, ISSN: 0163-6804. DOI: 10.1109/mcom.2014.6736761.

[48] A. F. Benzin, "Design and Implementation of centralized signal processing low latency massive MIMO SDR systems", Ph.D. dissertation, Technische Universität Berlin, Berlin, 2020. DOI: 10.14279/depositonce-10418.

[49] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO Achievable Rates With Downlink Training and Channel State Feedback", *IEEE Transactions on Information Theory*, vol. 56, no. 6, pp. 2845–2866, Jun. 2010, ISSN: 0018-9448. DOI: 10.1109/TIT.2010.2046225.

[50] J. Nam, J.-Y. Ahn, A. Adhikary, and G. Caire, "Joint spatial division and multiplexing: Realizing massive MIMO gains with limited channel state information", in *2012 46th Annual Conference on Information Sciences and Systems (CISS)*, Mar. 2012. DOI: 10.1109/CISS.2012.6310934.

[51] L. Miretti, R. L. G. Cavalcante, and S. Stanczak, "FDD Massive MIMO Channel Spatial Covariance Conversion Using Projection Methods", in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 3609–3613. DOI: 10.1109/ICASSP.2018.8462048.

[52] M. Barzegar Khalilsarai, S. Haghighatshoar, X. Yi, and G. Caire, "FDD Massive MIMO via UL/DL Channel Covariance Extrapolation and Active Channel Sparsification", *IEEE Transactions on Wireless Communications*, vol. 18, no. 1, pp. 121–135, Jan. 2019, ISSN: 1558-2248. DOI: 10.1109/TWC.2018.2877684.

[53] M. Shafi, A. F. Molisch, P. J. Smith, *et al.*, "5G: A Tutorial Overview of Standards, Trials, Challenges, Deployment, and Practice", *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 6, pp. 1201–1221, Jun. 2017, ISSN: 0733-8716. DOI: 10.1109/JSAC.2017.2692307.

[54] S. J. Orfanidis, *Electromagnetic Waves and Antennas*. Rutgers University, 2004. [Online]. Available: http://www.ece.rutgers.edu/~orfanidi/ewa/.

[55] D. Pozar, *Microwave Engineering*, Forth. Hoboken, NJ, USA: Wiley, Nov. 2011, ISBN: 9780470631553.

[56] T. S. Rappaport, G. R. MacCartney, S. Sun, H. Yan, and S. Deng, "Small-Scale, Local Area, and Transitional Millimeter Wave Propagation for 5G Communications", *IEEE Transactions on Antennas and Propagation*, vol. 65, no. 12, pp. 6474–6490, Dec. 2017. DOI: 10.1109/TAP.2017.2734159.

[57] IEEE Std 802.11ad-2012, "Wireless LAN medium access control (MAC) and physical layer (PHY) specifications AMENDMENT 3: Enhancements for very high throughput in the 60 GHz band", Tech. Rep., 2014. DOI: 10.1109/IEEESTD.2012.6392842.

[58] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter Wave Beamforming for Wireless Backhaul and Access in Small Cell Networks", *IEEE Transactions on Communications*, vol. 61, no. 10, pp. 4391–4403, Oct. 2013, ISSN: 0090-6778. DOI: 10.1109/TCOMM.2013.090513.120848.

[59] 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*, vol. 18, no. 3, pp. 1518–1533, 2019. DOI: 10.1109/TWC.2019.2892043.

[60] X. Lin, J. Li, R. Baldemair, *et al.*, "5G New Radio: Unveiling the Essentials of the Next Generation Wireless Access Technology", 2018. arXiv: 1806.06898 [cs.NI].

[61] G. Caire, "On the Ergodic Rate Lower Bounds With Applications to Massive MIMO", *IEEE Transactions on Wireless Communications*, vol. 17, no. 5, pp. 3258–3268, May 2018, ISSN: 1558-2248. DOI: 10.1109/TWC.2018.2808522.

[62] S. Mondal, R. Singh, A. I. Hussein, and J. Paramesh, "A 25-30 GHz Fully-Connected Hybrid Beamforming Receiver for MIMO Communication", *IEEE Journal of Solid-State Circuits*, vol. 53, no. 5, pp. 1275–1287, 2018. DOI: 10.1109/JSSC.2018.2789402.

[63] R. Rotman, M. Tur, and L. Yaron, "True Time Delay in Phased Arrays", *Proceedings of the IEEE*, vol. 104, no. 3, pp. 504–518, 2016. DOI: 10.1109/JPROC.2016.2515122.

[64] H. G. Myung, J. Lim, and D. J. Goodman, "Peak-To-Average Power Ratio of Single Carrier FDMA Signals with Pulse Shaping", in *2006 IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications*, Sep. 2006, pp. 1–5. DOI: 10.1109/PIMRC.2006.254407.

[65] S. Ju and T. S. Rappaport, "Millimeter-Wave Extended NYUSIM Channel Model for Spatial Consistency", in *2018 IEEE Global Communications Conference (GLOBECOM)*, Dec. 2018. DOI: 10.1109/GLOCOM.2018.8647188.

[66] 3GPP TR 38.901, "Study on channel model for frequencies from 0.5 to 100 GHz", Tech. Rep. v15.0.0, 2018. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3173.

[67] A. Benzin, G. Caire, Y. Shadmi, and A. M. Tulino, "Low-Complexity Truncated Polynomial Expansion DL Precoders and UL Receivers for Massive MIMO in Correlated Channels", *IEEE Transactions on Wireless Communications*, vol. 18, no. 2, pp. 1069–1084, 2019. DOI: 10.1109/TWC.2018.2889480.

[68] M. Agiwal, A. Roy, and N. Saxena, "Next Generation 5G Wireless Networks: A Comprehensive Survey", *IEEE Communications Surveys and Tutorials*, vol. 18, no. 3, pp. 1617–1655, 2016, ISSN: 1553-877X. DOI: 10.1109/COMST.2016.2532458.

[69] S. Jaeckel, L. Raschkowski, K. Börner, and L. Thiele, "QuaDRiGa: A 3-D Multi-Cell Channel Model With Time Evolution for Enabling Virtual Field Trials", *IEEE Transactions on Antennas and Propagation*, vol. 62, no. 6, pp. 3242–3256, Jun. 2014. DOI: 10.1109/TAP.2014.2310220.

[70] I. Ndip, K. Andersson, F. Dielacher, *et al.*, "Integration Specifications", SERENA, research rep., Jul. 2018. DOI: 10.5281/zenodo.3240451.

[71] T. Levanen, K. Pajukoski, M. Renfors, and M. Valkama, "Cost of Increased Bandwidth Efficiency in 5G NR", in *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, Sep. 2017, pp. 1–7. DOI: 10.1109/VTCFall.2017.8288107.

[72] P. Serrano, P. Salvador, V. Mancuso, and Y. Grunenberger, "Experimenting With Commodity 802.11 Hardware: Overview and Future Directions", *IEEE Communications Surveys and Tutorials*, vol. 17, no. 2, pp. 671–699, 2015. DOI: 10.1109/comst.2015.2417493.

[73] Y. Xie, Y. Zhang, J. C. Liando, and M. Li, "SWAN: Stitched Wi-Fi ANtennas", in *24th ACM International Conference on Mobile Computing and Networking (MobiCom 2018)*, New Delhi, India, Oct. 2018, pp. 51–66, ISBN: 978-1-4503-5903-0. DOI: 10.1145/3241539.3241572.

[74] V. Arun and H. Balakrishnan, "RFocus: Practical Beamforming for Small Devices", 2019. arXiv: 1905.05130 [cs.NI].

[75] M. Dunna, C. Zhang, D. Sievenpiper, and D. Bharadia, "ScatterMIMO: enabling virtual MIMO with smart surfaces", in *26th ACM International Conference on Mobile Computing and Networking (MobiCom 2020)*, London, United Kingdom, Sep. 2020, ISBN: 9781450370851. DOI: 10.1145/3372224.3380887.

[76] D. Bharadia and S. Katti, "FastForward: Fast and Constructive Full Duplex Relays", *ACM SIGCOMM Computer Communication Review*, vol. 44, no. 4, pp. 199–210, Aug. 2014, ISSN: 0146-4833. DOI: 10.1145/2740070.2626327.

[77] E. Hamed, H. Rahul, M. A. Abdelghany, and D. Katabi, "Real-Time Distributed MIMO Systems", in *Proceedings of the 2016 ACM SIGCOMM Conference*, Florianopolis, Brazil, 2016, pp. 412–425. DOI: 10.1145/2934872.2934905.

[78] Z. Li, Y. Xie, L. Shangguan, *et al.*, "Towards Programming the Radio Environment with Large Arrays of Inexpensive Antennas", in *16th USENIX Symposium on Networked Systems Design and Implementation (NSDI 19)*, Boston, MA, USA, Feb. 2019, pp. 285–300, ISBN: 978-1-931971-49-2. [Online]. Available: https://www.usenix.org/conference/nsdi19/presentation/lizhuqi.

[79] J. Gjengset, J. Xiong, G. McPhillips, and K. Jamieson, "Phaser: enabling phased array signal processing on commodity WiFi access points", in *20th ACM International Conference on Mobile Computing and Networking (MobiCom 2014)*, Maui, Hawaii, USA, Sep. 2014, pp. 153–164, ISBN: 978-1-4503-2783-1. DOI: 10.1145/2639108.2639139.