D1.4

Signal processing algorithms and specifications

Project number:	779305
Project acronym:	**SERENA**
Project title:	Gan-on-Silicon Efficient mm-Wave European System Integration Platform
Project Start Date:	1st January, 2018
Duration:	36 months
Programme:	H2020-ICT-2017-1
Deliverable Type:	Report
Reference Number:	ICT-779305 / D1.4 / 1.0
Workpackage:	WP1
Due Date:	December 2018 – M12
Actual Submission Date:	21st December, 2018
Responsible Organisation:	TUB
Editor:	Thomas Kühne
Dissemination Level:	PU
Revision:	1.0
Abstract:	This deliverable describes the signal processing of the SERENA 5G 39 GHz proof-of-concept demonstrator. The report introduces different signal processing aspects of the system including the hybrid digital-analog architecture, the initial acquisition phase, and the multiuser multiple-input multiple-output precoding. It proposes solutions for these problems and describes them theoretically.
Keywords:	Specification, signal processing, hybrid digital-analog beamforming, multiuser multiple-input multiple-output

The SERENA project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 779305.

Editor

Thomas Kühne (TUB)

Contributors (ordered according to beneficiary numbers)

Thomas Kühne (TUB)
Giuseppe Caire (TUB)
Xiaoshen Song (TUB)

Reviewers

Kristoffer Andersson (EAB)
Ulf Gustafsson (EAB)
Kimmo Rasilainen (Chalmers)

Disclaimer

Executive Summary

This report gives an introduction to the signal processing of hybrid digital-analog systems like the SERENA platform. It focuses on the parts of the signal processing, which are influenced by the hybrid structure. These are the architecture of the hybrid digital-analog structure, the initial acquisition, and the multiuser multiple-input multiple-output precoding. These parts are necessary to demonstrate the 5G scenario of the SERENA proof-of-concept platform. The report proposes algorithms and shows simulation results of their performance.
The proof-of-concept platform will use a sub-connected hybrid digital-analog architecture. A comparison with an alternative fully-connected architecture analyzes their power efficiency and communication performance. The chosen architecture does not have the same performance as the alternative but has a much lower complexity, which justifies its use. A proposed advanced algorithm for the initial acquisition phase outperforms the literature. Numerical simulations show its performance. Additional basic schemes are described and used as a benchmark. A precoding algorithm for the data communication phase is introduced. Through simulations, the spectral efficiency is characterized.
Every part of the report summarizes the decision for the signal processing of the proof-of-concept system. Common signal processing parts which are not explained are given in the introduction. They are based on common standards. The chosen algorithms will enable the 5G scenario of the SERENA platform and from a theoretical point of view achieve the corresponding goals of the project. The deliverable estimates an increase by a factor of 29 to 229 for the wireless area capacity compared to LTE advanced depending on the cell size.

Contents

List of Figures

Chapter 1

Introduction

1.1 General considerations

This report summarizes the investigation of the theoretical signal processing aspects of the SERENA project. It presents different key parts of the signal processing necessary for the 39 GHz 5G proof-of-concept demonstrator. The presented algorithms and specifications will be implemented in WP6 to show and measure the performance of the proof-of-concept system.
A major challenge of the communication at mm-waves is the very high free-space path loss. According to Friis transmission formula, the effective area of an isotropic antenna decreases polynomially with frequency. Therefore, the free-space path loss at mm-waves is considerably larger compared to sub-6 GHz systems [1]. The SERENA project tries to counteract the increased path loss with two solutions. First, by increasing the available output power by using the high power Gallium Nitride technology for the power amplifiers (PAs). Secondly, by using a large antenna array especially at the base station. The small wavelength of mm-wave signals enables to pack a large number of antenna elements in a small form factor. The antenna array provides a large beamforming gain to achieve the required link budget. Additionally, a large antenna array enables the use of multiuser multiple-input multiple-output (MU-MIMO) precoding to increase the data rate of the base station.
To obtain the beamforming and MU-MIMO gain, the phase and amplitude of the signals of all antenna elements need to be independently controllable. In general, the associated signal processing is called precoding at the transmitter (TX) and detection at the receiver (RX). It can be implemented either in the digital baseband on a signal processing platform or with a controllable analog network in the radio frequency (RF) signal paths. In the digital case, every antenna element is connected via an RF chain (including frequency converters, analog to digital converters, etc.) to a digital baseband signal. Conventional full-digital baseband precoding for large bandwidth signals with many antenna elements results in an unaffordable hardware cost. Instead, a combination of both digital and analog precoding, using a reduced number of RF chains to convert the signals to the digital baseband, known as hybrid digital-analog (HDA) structure, has been widely considered [2] and is used in the SERENA project.
The scope of this report and the described work is to investigate the required signal processing to use the advancements of the hardware concept of the SERENA project to maximize the performance gain. The HDA approach leads to three major challenges in the signal processing compared to full-digital MU-MIMO precoding or to non-MU-MIMO signal processing. First, the HDA architecture with the partitioning between the digital and analog processing needs to be examined and defined. This is presented in Chapter 2. Second, the initial acquisition phase of the communication, where the users are detected by the base station and the channel state

information (CSI) is acquired, is challenging due to the low number of RF chains respectively measurements compared to the number of antenna elements. We refer to the initial acquisition in mm-wave systems as beam alignment (BA) and describe it in Chapter 3. Third, the precoding algorithms for the data transmission have to be adapted to the special limitations which the HDA architecture has. We will compare possible precoding algorithms in Chapter 4.
Other aspects of the signal processing were not explicitly investigated since they do not deviate from sub-6 GHz 5G systems. Nonetheless, the final proof-of-concept platform is planned to demonstrate the communication performance. For this, the system will be as close as possible and necessary to the 5G standard. As modulation, we will use orthogonal frequency division multiplexing (OFDM) with quadrature amplitude modulation (QAM). The system will operate in time division multiplexing (TDD) mode as it is necessary to preserve the MU-MIMO gain with a large number of antennas and users [3].

1.2 Channel model

For the mathematical investigation of the BA phase (Chapter 3) and the data communication phase (Chapter 4) we use the following channel model. We consider a system formed by a base station equipped with a uniform linear array (ULA) with N_{BS} antennas and N_{RF} RF chains serving simultaneously $K = N_{\mathsf{RF}}$ user equipments (UEs), each of which is also equipped with a ULA with N_{UE} antennas and one RF chain. The propagation channel between the base station and the k-th UE, $k \in [K]$, consists of $L_k \ll \max\{D, N_{\mathsf{UE}}\}$ multi-path components, where the baseband equivalent impulse response of the channel at time slot s reads

$$\begin{aligned}\mathsf{H}_{k,s}(t,\tau) &= \sum_{l=1}^{L_k} \rho_{k,s,l} e^{j2\pi\nu_{k,l}t} \mathbf{a}_{\mathsf{R}}(\phi_{k,l})\mathbf{a}_{\mathsf{T}}(\theta_{k,l})^{\mathsf{H}} \delta(\tau-\tau_{k,l}) \\ &= \sum_{l=1}^{L_k} \mathsf{H}_{k,s,l}(t)\delta(\tau-\tau_{k,l}), \end{aligned} \tag{1.1}$$

where $\mathsf{H}_{k,s,l}(t) := \rho_{k,s,l} e^{j2\pi\nu_{k,l}t} \mathbf{a}_{\mathsf{R}}(\phi_{k,l})\mathbf{a}_{\mathsf{T}}(\theta_{k,l})^{\mathsf{H}}$, $(\phi_{k,l}, \theta_{k,l}, \tau_{k,l}, \nu_{k,l})$ denote the angle of arrival (AoA), angle of departure (AoD), delay, and Doppler shift of the l-th component, and $\delta(\cdot)$ denotes the Dirac delta function. The vectors $\mathbf{a}_{\mathsf{T}}(\theta_{k,l}) \in \mathbb{C}^D$ and $\mathbf{a}_{\mathsf{R}}(\phi_{k,l}) \in \mathbb{C}^{N_{\mathsf{UE}}}$ are the array response vectors of the base station and UE at AoD $\theta_{k,l}$ and AoA $\phi_{k,l}$ respectively, with elements given by

$$[\mathbf{a}_{\mathsf{T}}(\theta)]_d = e^{j(d-1)\pi\sin(\theta)},\ d \in [D], \tag{1.2a}$$

$$[\mathbf{a}_{\mathsf{R}}(\phi)]_n = e^{j(n-1)\pi\sin(\phi)},\ n \in [N_{\mathsf{UE}}], \tag{1.2b}$$

Here we assume that the spacing of the ULA antennas equals to the half of the wavelength. Also note, that we do use D instead of N_{BS} because of the "effective" number of elements at the base station depends on the hybrid architecture, which will be explained in the following chapter. We adopt a block fading model, i.e., the channel gains $\rho_{k,s,l}$ remain invariant over the channel *coherence time* Δt_c but change independent and identically distributed randomly across different Δt_c. Since each scatterer in practice is a superposition of many smaller components that have (roughly) the same AoA-AoD and delay, we assume a general Rice fading model given by

$$\rho_{k,s,l} \sim \sqrt{\gamma_{k,l}} \left(\sqrt{\frac{\eta_{k,l}}{1+\eta_{k,l}}} + \frac{1}{\sqrt{1+\eta_{k,l}}} \check{\rho}_{k,s,l} \right), \tag{1.3}$$

where $\gamma_{k,l}$ denotes the overall multi-path component strength, $\eta_{k,l} \in [0, \infty)$ indicates the strength ratio between the line-of-sight and the non-line-of-sight components, and $\check{\rho}_{k,s,l} \sim \mathcal{CN}(0,1)$ is a zero-mean unit-variance complex Gaussian random variable. In particular, $\eta_{k,l} \to \infty$ indicates a pure line-of-sight path while $\eta_{k,l} = 0$ indicates a pure non-line-of-sight path, affected by standard Rayleigh fading.

Following the *beamspace representation* as in [4], we obtain an approximate finite-dimensional representation of the channel response (1.1) with respect to the discrete dictionary in the AoA-AoD (beam) domain defined by the quantized angles

$$\Phi := \{\check{\phi} : (1+\sin(\check{\phi}))/2 = \frac{n-1}{N_{\mathsf{UE}}},\, n \in [N_{\mathsf{UE}}]\}, \tag{1.4a}$$

$$\Theta := \{\check{\theta} : (1+\sin(\check{\theta}))/2 = \frac{d-1}{N_{\mathsf{BS}}}, d \in [N_{\mathsf{BS}}]\}, \tag{1.4b}$$

with corresponding array response vectors $\mathcal{A}_{\mathsf{R}} := \{\mathbf{a}_{\mathsf{R}}(\check{\phi}) : \check{\phi} \in \Phi\}$ and $\mathcal{A}_{\mathsf{T}} := \{\mathbf{a}_{\mathsf{T}}(\check{\theta}) : \check{\theta} \in \Theta\}$. For ULAs as considered in this paper, the dictionaries $\mathcal{A}_{\mathsf{R}}$ and $\mathcal{A}_{\mathsf{T}}$, after suitable normalization, yield to the discrete Fourier transform (DFT) matrices $\mathbf{F}_{\mathsf{UE}} \in \mathbb{C}^{N_{\mathsf{UE}} \times N_{\mathsf{UE}}}$ and $\mathbf{F}_{\mathsf{BS}} \in \mathbb{C}^{D \times N_{\mathsf{BS}}}$ with elements

$$[\mathbf{F}_{\mathsf{UE}}]_{n,n'} = \frac{1}{\sqrt{N_{\mathsf{UE}}}} e^{j2\pi(n-1)(\frac{n'-1}{N_{\mathsf{UE}}}-\frac{1}{2})}, n, n' \in [N_{\mathsf{UE}}], \tag{1.5a}$$

$$[\mathbf{F}_{\mathsf{BS}}]_{d,d'} = \frac{1}{\sqrt{N_{\mathsf{BS}}}} e^{j2\pi(d-1)(\frac{d'-1}{N_{\mathsf{BS}}}-\frac{1}{2})}, d \in [D], d' \in [N_{\mathsf{BS}}]. \tag{1.5b}$$

Consequently, the beam-domain channel representation reads

$$\check{\mathsf{H}}_{k,s}(t,\tau) = \mathbf{F}_{\mathsf{UE}}^{\mathsf{H}} \mathsf{H}_{k,s}(t,\tau) \mathbf{F}_{\mathsf{BS}} = \sum_{l=1}^{L_k} \check{\mathsf{H}}_{k,s,l}(t) \delta(\tau - \tau_{k,l}), \tag{1.6}$$

where $\check{\mathsf{H}}_{k,s,l}(t) := \mathbf{F}_{\mathsf{UE}}^{\mathsf{H}} \mathsf{H}_{k,s,l}(t) \mathbf{F}_{\mathsf{BS}}$. It is well-known (e.g., see [5] and references therein) that, as N_{BS} and N_{UE} increase, the DFT basis provides a very sparse channel representation.

Chapter 2

Hybrid digital-analog architecture

2.1 Description of the architectures

The HDA approach as a combination of digital and analog signal processing consists of an analog network which connects a large number of antennas N_{BS} to a smaller number of RF chains N_{RF} and a digital signal processing part. The HDA architecture is defined by this reduction $N_{\text{BS}}/N_{\text{RF}}$ and its implementation. For simplicity, we consider the transmission case in the following mathematical modeling. The reduction is realized through the architecture of the analog network and can mathematically be expressed as matrix $\tilde{\mathbf{U}} \in \mathbb{C}^{N_{\text{RF}} \times N_{\text{BS}}}$. We denote the signals of the RF chains as $\mathbf{x} = [x_1, \cdots, x_{N_{\text{RF}}}]$. Following, the beamformed output can be written as

$$\tilde{\mathbf{x}} = \sqrt{\alpha_{\text{com}}}\tilde{\mathbf{U}}\sqrt{\alpha_{\text{div}}}\mathbf{x} \tag{2.1}$$

where α_{com} and α_{div} represent the power combiners and dividers in the network, respectively. 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. The type of the controllable component can either be an on-off switch, a phase shifter, or a vector modulator (with a variable amplitude and phase). The elements of $\tilde{\mathbf{U}}$ are limited by an on-off switch to be $[\tilde{\mathbf{U}}] \in \{0, 1\}$, by a phase shifter to be $[\tilde{\mathbf{U}}]_{l,k} = e^{j\phi_{l,k}}$, and by a vector modulator to be $[\tilde{\mathbf{U}}]_{l,k} = a_{l,k}e^{j\phi_{r,c}}$. Here $\phi_{l,k}$, and $a_{l,k}$ are the selected phase and amplitude, again. Therefore, mathematically the on-off switch is a subset of the phase shifter which is a subset of the vector modulator. Since the beamforming integrated circuit (IC) of the SERENA platform supports phase and amplitude control we do only consider the vector modulator case in the following investigation. Vector modulators enable the highest degree of freedom in the algorithms but are more complex to implement than the other options. In the literature, many structures of the network interconnections have been proposed (see [6]). The two main network types are the fully-connected (FC) and the one-stream-per-subarray (OSPS) structure. The FC architecture is shown in Fig. 2.1(a). Every RF chain is connected to every antenna element. Hence, $\tilde{\mathbf{U}}$ is a full matrix with the form $[\tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_2, \cdots, \tilde{\mathbf{u}}_{N_{\text{RF}}}]$ where the vector $\tilde{\mathbf{u}}_k \in \mathbb{C}^{N_{\text{BS}}}$ is the beamforming vector for the k-th RF chain. In the FC case, $\alpha_{\text{com}} = 1/N_{\text{RF}}$ and $\alpha_{\text{div}} = 1/N_{\text{BS}}$. Fig. 2.1(b) shows the OSPS architecture. In this architecture, 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 $\tilde{\mathbf{U}}$ having the form

$$\begin{bmatrix} \tilde{\mathbf{u}}_1 & 0 & \dots & 0 \\ 0 & \tilde{\mathbf{u}}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \tilde{\mathbf{u}}_{N_{\text{RF}}} \end{bmatrix} \tag{2.2}$$

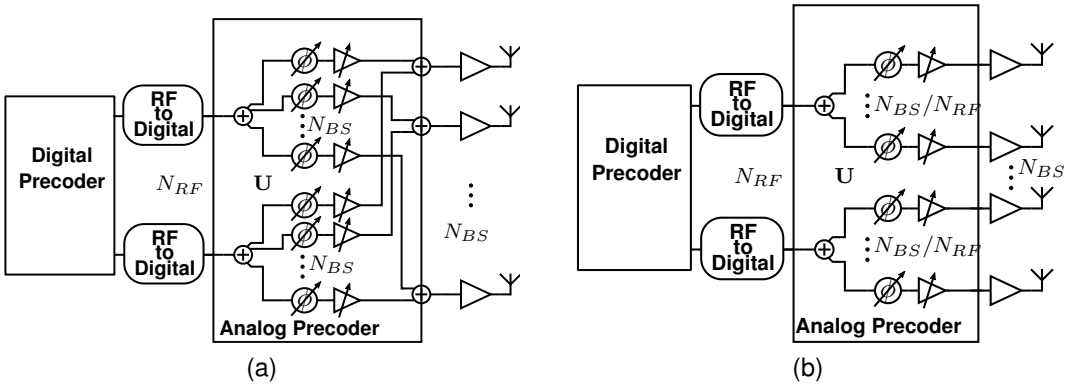

Figure 2.1: Hybrid transmitter architectures: (a) fully-connected (FC), and (b) one-stream-per-subarray (OSPS).

where $\tilde{\mathbf{u}}_k \in \mathbb{C}^{N_{\text{BS}}/N_{\text{RF}}}$ is again the beamforming vector for the k-th RF chain but with the size $N_{\text{BS}}/N_{\text{RF}}$. In the OSPS case, $\alpha_{\text{com}} = 1$ and $\alpha_{\text{div}} = N_{\text{BS}}/N_{\text{RF}}$. In the OSPS architecture are only $N_{\text{BS}}/N_{\text{RF}}$ antenna elements connected to an RF chain. In contrast, in the FC architecture all N_{BS} elements are connected to an RF chain. Therefore, the narrowest beams of the OSPS structure in the angular domain are wider than in the FC structure. Hence, in the channel model the size of the array response defined in equation (1.2a) and following is $D = N_{\text{BS}}/N_{\text{RF}}$ for the OSPS structure whereas for the FC architecture it is $D = N_{\text{BS}}$.
Comparing both architectures, an FC network has a full matrix $\tilde{\mathbf{U}}$ but it becomes very complex for large arrays and many RF chains as it requires $N_{\text{RF}} \times N_{\text{BS}}$ controllable components. The FC architecture is infeasible with state-of-the-art technology for mm-wave systems. The size and the cost of such a system is very high. In addition, as the size of the network becomes large, the insertion loss of the interconnections is a problem. On the other hand, the OSPS structure reduces greatly the complexity of the hardware implementation but at the cost of a less adjustable matrix $\tilde{\mathbf{U}}$. This might limit the BA and precoding performance of the system. Performance comparisons of both architectures are presented in the following chapters. Additionally, an FC network requires more power dividers/combiners than an OSPS network resulting in smaller factors α_{div} and α_{com}. These factors represent losses in the system and need to be balanced with amplification to achieve the same output power.
In SERENA, the architecture is predefined by both the beamforming IC, which has a one to four network between the RF ports, and the integration module, which is limited to one beamformer implying an OSPS structure. Since both the beamformer and the integration are already beyond state-of-the-art these limits are fixed. Therefore the SERENA platform will have an OSPS architecture with a minimum ration of $N_{\text{BS}}/N_{\text{RF}} = 4$. To further reduce the number of RF chains and therefore the implementation cost this ratio could be increased by combining the signals from multiple integration modules. For the proof-of-concept system, this ratio will most likely not be increased since the TUB CommIT massive MIMO platform and the SERENA system board provide enough RF chains. To verify theoretical results with experimentation on the proof-of-concept system, a larger ratio can be simulated by restricting the digital signal processing part.

2.2 Power efficiency modeling

In addition to the BA and communication performance, also the power efficiency depends on the architecture of the analog network. 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 architecture 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 digital signal processing might not be negligible but it does not depend on the architecture and can therefore also be ignored.
The sum-power of the vector $\tilde{\mathbf{x}}$ of the beamformed signals, which are the inputs for the PAs, is

$$\begin{aligned}\tilde{P} &= \mathbb{E}[\tilde{\mathbf{x}}^{\mathsf{H}}\tilde{\mathbf{x}}] = \alpha_{\mathsf{com}}\alpha_{\mathsf{div}} \cdot \mathbf{x}^{\mathsf{H}}\tilde{\mathbf{U}}^{\mathsf{H}}\tilde{\mathbf{U}}\mathbf{x} \\ &= \alpha_{\mathsf{com}}\alpha_{\mathsf{div}} \cdot \mathsf{tr}\left(\mathbf{x}\mathbf{x}^{\mathsf{H}}\tilde{\mathbf{U}}^{\mathsf{H}}\tilde{\mathbf{U}}\right)\end{aligned} \tag{2.3}$$

where $\tilde{\mathbf{U}}^{\mathsf{H}}$ is the conjugate transpose of $\tilde{\mathbf{U}}$ and $\mathtt{tr}()$ represents the trace of a matrix. Since we do not consider the power consumption of the amplification within the network or generating the input power to the network we can set both α_{div} and α_{com} to 1.
The beamformed signal is amplified at every antenna element by a PA. All PAs have the same amplification, maximum output power P_{max}, and efficiency at maximum power η_{max}. For any given element in the array, let P_{rad} denote the radiated power of the element, and P_{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 [7] as

$$P_{\mathsf{cons}} = \frac{\sqrt{P_{\mathsf{max}}}}{\eta_{\mathsf{max}}}\sqrt{P_{\mathsf{rad}}}. \tag{2.4}$$

This function 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

$$\eta_{\mathsf{eff}} = \frac{P_{\mathsf{rad}}}{P_{\mathsf{cons}}}. \tag{2.5}$$

Due to the beamforming, the superposition of multiple beamforming vectors (particularly in the FC case), and the potentially high peak-to-average power ratio (PAPR) of the time-domain transmit waveform x_k (particularly with an OFDM modulation), the input power of the N_{BS} PAs varies. To avoid non-linear distortion the system needs to assure that every PA works below its maximum output power. To compare the two transmitter architectures with different precoding/beamforming strategies and different modulation schemes, we generally have two options:
Option I: Both architectures utilize the same PA but apply a different input backoff $\alpha_{\mathsf{off}} \in (0, 1]$, such that the peak power of the radiated signal is smaller than P_{max}. As a reference, we denote as $(P_{\mathsf{rad},0}, \eta_{\mathsf{max},0})$ the parameters of this reference PA under the reference precoding/beamforming strategy with a power backoff factor $\alpha_{\mathsf{off},0}$ (as illustrated later in this chapter). For different signal processing scenarios (with certain α_{off}) the effective radiated power and the consumed power read $P_{\mathsf{rad}} = \frac{\alpha_{\mathsf{off}}}{\alpha_{\mathsf{off},0}}P_{\mathsf{rad},0}$ and $P_{\mathsf{cons}} = \frac{\sqrt{P_{\mathsf{max},0}}}{\eta_{\mathsf{max},0}}\sqrt{P_{\mathsf{rad}}}$. The transmitter efficiency is given by

$$\eta_{\mathsf{eff}} = \frac{P_{\mathsf{rad}}}{P_{\mathsf{cons}}} = \frac{\sqrt{P_{\mathsf{rad}}} \cdot \eta_{\mathsf{max},0}}{\sqrt{P_{\mathsf{max},0}}}. \tag{2.6}$$

Option II: We choose to deploy different PAs for different architectures, with a maximum output power given by $P_{\max} = \frac{\alpha_{\mathsf{off},0}}{\alpha_{\mathsf{off}}}P_{\mathsf{max},0}$, where α_{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 $P_{\text{cons}} = \frac{\sqrt{P_{\text{max},0} \cdot \alpha_{\text{off},0}/\alpha_{\text{off}}}}{\eta_{\text{max}}} \sqrt{P_{\text{rad}}}$. The transmitter efficiency is given by

$$\eta_{\text{eff}} = \frac{P_{\text{rad}}}{P_{\text{cons}}} = \frac{\sqrt{P_{\text{rad}}} \cdot \eta_{\text{max}}}{\sqrt{P_{\text{max},0} \cdot \alpha_{\text{off},0}}} \cdot \sqrt{\alpha_{\text{off}}}. \tag{2.7}$$

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

2.3 Power efficiency simulations

We have run numerical simulations to compare both architectures. We also simulated two modulation schemes, single-carrier (SC) and OFDM modulation. The PAPR and therefore the efficiency with a linear operation of the PAs depends on the modulation and the architecture. We consider both options from the previous section and a system with $N_{\text{BS}} = 64$ base station antenna elements and $N_{\text{RF}} = 8$ RF chains.

We first assume a reference scenario as the baseline, i.e the OSPS architecture using an SC modulation. The reference PA has $P_{\text{max},0} = 36$ dBm and $\eta_{\text{max},0} = 0.24$. The backoff factor with respect to different waveforms and transmitter architectures can be written as $\alpha_{\text{off}} = 1/(P_{\text{PAPR}})$, where P_{PAPR} represents the PAPR of the input signals at the PAs. The investigation for 3GPP LTE in [8] showed that with a probability of 0.9999, the PAPR of the LTE SC 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_{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 [8] 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 SC signaling there is no clear work in the literature that shows how the sum of SC signals behaves. We simulated the sum of $N_{\text{RF}} = 8$ SC signals using the same parameters as in [8]. The result shows that with a probability of 0.9999, the PAPR of the sum is smaller than ~ 11.0 dB. We apply these values and, without loss of generality, we assume $\alpha_{\text{off},0} = 0$ dB as reference.

As can be seen in Equation (2.6), the efficiency with the assumptions of *option I* only depends on the chosen PA and the radiated power P_{rad}. P_{rad} itself does not depend directly on the architecture but on the necessary backoff α_{off}. Fig. 2.2(a) shows the achievable radiated power for the different architectures and modulation schemes versus the radiated power of the reference scenario (OSPS structure and SC modulation). The OSPS structure with an SC modulation results in the highest radiated power followed by the FC structure with an SC modulation. For the OFDM modulation, both architectures have the same efficiency and radiated power which are smaller than the SC modulation values. Fig. 2.2(b) shows the efficiency against the radiated power in the case of *option II*. The difference between the architectures and modulation schemes can be explained with 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 SC modulation has the highest efficiency, followed by the FC structure with SC modulation. Both architectures with an OFDM modulation have the lowest efficiency.

A conclusion of the modeling and the simulations is, that the architecture does not directly influence the efficiency of the system (when only considering the PAs). Rather, the architecture can change the PAPR of the beamformed signal which in turn impacts the efficiency.

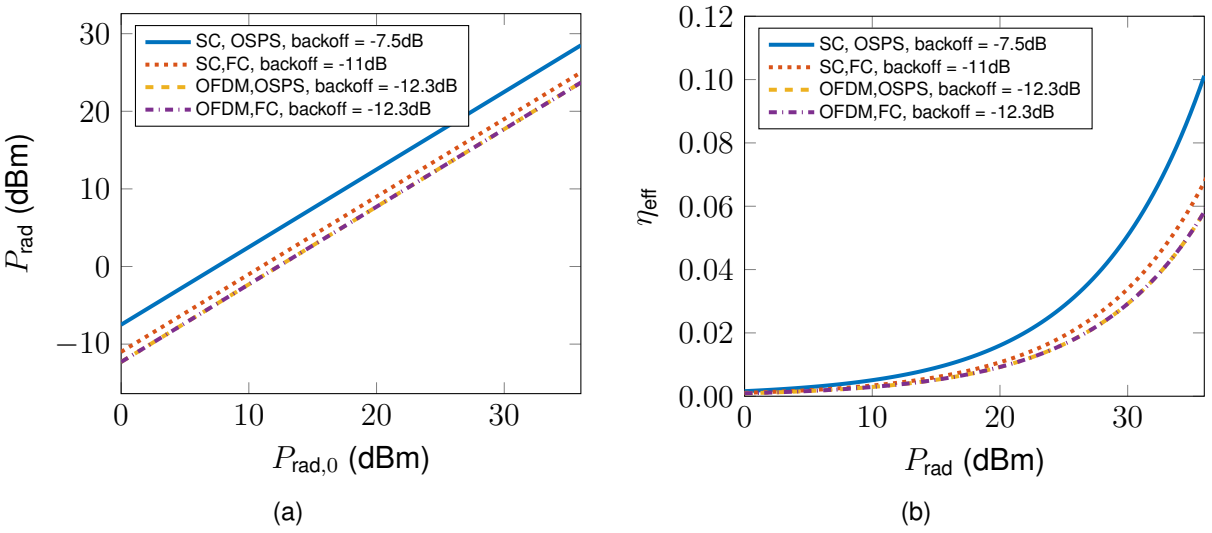

Figure 2.2: The power efficiency evaluation of both architectures in terms of (a) the actual radiated power in *option I* vs. the radiated power of the reference scenario, and (b) the power efficiency in *option II* vs. the actual radiated power.

Chapter 3

Beam alignment phase

The initial phase in a cellular communication process is the acquisition of users by the base station. This includes the measurement of the CSI. Compared to sub-6 GHz full-digital systems a mm-wave hybrid system suffers from two main problems. First, the high path loss prohibits the direct omnidirectional measurement of the channel. In a mm-wave system beamforming at both sides is needed in order to provide a sufficient signal-to-noise ratio (SNR). Hence, both the base station and the UE must train their beams to point at an angle of arrival (AoA)/angle of departure (AoD) corresponding to a multipath component that conveys enough signal energy. Therefore, we refer to the initial acquisition in millimeter-wave systems as beam alignment (BA). The second problem of a hybrid mm-wave system compared to a full-digital system is the low number of RF chains. Because of the large number of antenna elements at both sides, the size of the channel matrix between each UE and the base station is very large. The low number of RF chains restricts the measurement rate of the large channel. The base station measures per training slot respectively orthogonal resource[1] only a low-dimensional representation of the large channel matrix. Depending on the algorithm, the system uses multiple training slots to estimate the channel or a certain representation of it. The overall time necessary for the BA can become very large and limit the sum rate of the system. However, 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 [9]. As a result, a suitable BA scheme only needs to identify a very small subset of the channel. Hence, the BA algorithm needs to be optimized to achieve an optimal system performance.
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 [10] (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. We introduce an advanced algorithm in Section 3.2.

3.1 Basic algorithms

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 phased array with optimal phase coefficients [11, p. 1088] to form narrow beams with a certain beam direction and angle and with a given beamwidth $\Delta\theta$. $\Delta\theta$ is dependent on N_{BS} and the direction of the beam. The antenna directivity (equally the array factor, if we ignore the single element pattern) generated with the

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

phased array is equivalent to the maximum possible directivity using a set number of elements. A codebook $\mathcal{C}_{\mathsf{BS}}$ of beamforming vectors for the base station is formed to cover the angular interval with N_{BS} equidistant beams. The UE forms an equivalent codebook $\mathcal{C}_{\mathsf{UE}}$. Since both the base station 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 $\mathcal{C}_{\mathsf{BS}}$ and $\mathcal{C}_{\mathsf{UE}}$ and stores the corresponding channel gain. If the search requires less than the coherence time of the channel it measures the instantaneous full CSI. The exhaustive search requires $N_{\mathsf{BS}} \times N_{\mathsf{UE}}$ training slots. For a simulation-based analysis of the exhaustive search algorithm, see [12].
Another basic algorithm is a hierarchical bisection search for example described in [13]. The idea of a hierarchical bisection search is to use multiple codebooks in levels. A codebook of level i is divided into J^{i-1} subsets with J beamforming vectors in each subset j. $\mathcal{C}_{i,j}$ is a codebook of level $i = 1, 2...I$ and subset $j = 1, 2...J^{i-1}$, where I is the number of levels and J the number of beams per subset. One subset of a subsequent level is a refined version of one codeword respectively beam of the previous level, covering with J beams the angular interval covered by the one previous beam. A typical choice $J = 2$ means that the angular interval of a beam is divided into 2 beams in the next level. $[C_{(i,j)}]_{:,m}$ is the beamforming vector $m = 1, 2..J$ in level i and subset j. The highest level has N_{BS} beam vectors with the same angular resolution as the exhaustive search. The beamforming gain of codewords increases with the level of the codebook. Therefore, a lower level has a higher probability of error for finding the correct beam covering the AoA. This effect could be compensated, for example, by using multiple or longer training slots. Different levels would require a different training time. However, this would increase the training time and therefore goes against the motivation for the algorithm.
Different approaches can be used to realize the beamforming vectors. Reference [13] proposes a version of an orthogonal matching pursuit algorithm to calculate the beamforming coefficients. Another way to generate the lower levels is with a quadratic phase excitation as described in [14]. The latter creates less ideal beams but requires only one RF chain per codeword. Whereas the former requires a large number of RF chains. The search process iterates through the levels of the codebook, refining the beamforming. Similar to the exhaustive search algorithm, in case of antenna arrays at the base station and the UEs the hierarchical bisection search needs a form of synchronization and a feedback channel between the UE and the base station. Assuming synchronization, the system measures in the current level / subset every combination of codewords of $\mathcal{C}^{\mathsf{BS}}_{i,j}$ and $\mathcal{C}^{\mathsf{UE}}_{i,j}$. If the channel is measured in the downlink by the UE, it feeds the strongest beam of the measured subset back to the base station. In every subsequent level, both the base station and UE use the subsets of the beams corresponding to the highest channel gain. To estimate the strongest path between a user and a base station the hierarchical search needs $J^2 \log_J(N_{\mathsf{UE}}) + J \log_J(N_{\mathsf{BS}}/N_{\mathsf{UE}})$ training slots. As described, the algorithm does not use available RF chains. One could imagine that the J beams per subset could be measured at the same time using multiple RF chains.
We have tested multiple basic algorithms including the both mentioned above with measurements using an HDA platform at $2.4\,\mathrm{GHz}$. Due to the different frequency and system specifications, the exact results cannot directly be transferred to the SERENA specifications. Nevertheless, the purpose of the measurements is to give some general insights that are still valid for SERENA. The measurements were conducted in a large empty hall with a ground area of $20\,\mathrm{m}$ by $7\,\mathrm{m}$ and a height of $6\,\mathrm{m}$. The base station had $N_{\mathsf{BS}} = 16$ antenna elements and $N_{\mathsf{RF}} = 2$ RF chains. The UE used a single fixed-beam antenna. This setup simulated a sparse environment with a strong line-of-sight path. The system measured the AoA at the base station for a training sequence sent in uplink by the UE. In addition to the exhaustive search and the hierarchical bisection search we

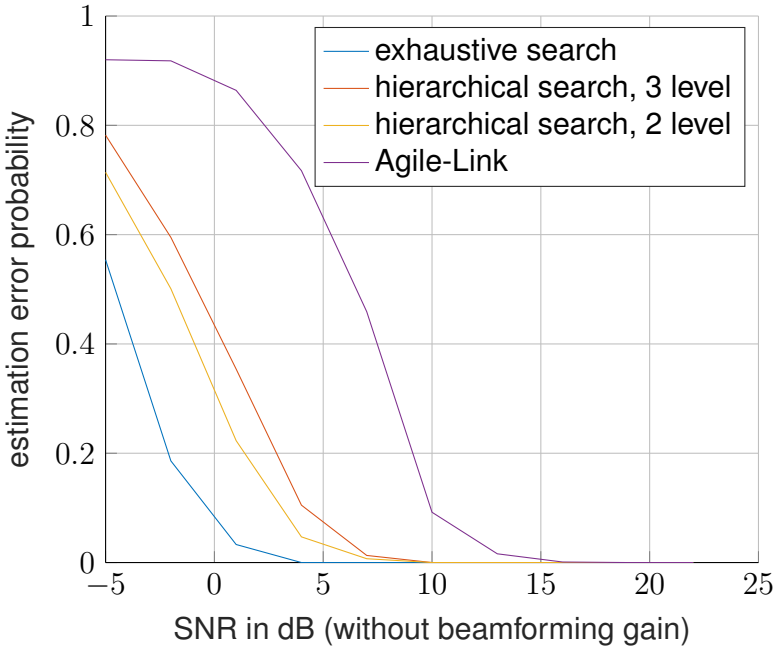

Figure 3.1: Measured average probability of error estimating the AoA for the basic BA algorithms.

have also measured an algorithm called Agile-Link proposed in [15]. We measured two versions of the hierarchical bisection search. A 3-level version as described above and a 2-level version. In the 2-level case, we skipped the first level of the algorithm. This version directly measures all codewords of the second level. This does not increase the number of necessary training slots but, due to the start with narrower beams, increases the beamforming gain of the first measured level.

Fig. 3.1 shows the estimation error probability versus the SNR without beamforming gain. The exhaustive search shows the best performance. The 2-level hierarchical search algorithm has a gap of around $3\,\mathrm{dB}$ and the 3-level configuration $5\,\mathrm{dB}$ compared to the exhaustive search. The gap between Agile-Link and the exhaustive search is $11\,\mathrm{dB}$. Taking measurement tolerances into account, the size of the gaps is roughly equal to the difference between the lowest beamforming gain of the codebooks and the gain of the exhaustive search beams. We can conclude that for the basic algorithms, the error probability performance is mainly determined by the lowest beamforming gain of all codewords. Additional information on the measurements can be found in our publication [16].

3.2 Advanced algorithm

In contrast to the basic schemes described above, one goal of more advanced BA algorithms is to use a minimum of training resources. Due to the sparse nature of mm-wave channels, compressed sensing is considered as a powerful technique to reduce the number of training slots. There are different algorithms proposed in the literature (e.g. [17, 18]), including our own [5, 4]. The approach of [5] is for OFDM systems, whereas [4] uses pseudo-noise sequences in the time-domain as training resource. Both versions of our algorithm estimate the second-order statistics of the channel. The coherence time of the second-order statistics is much larger than that of the instantaneous CSI. This makes our scheme robust to large Doppler spreads and fast channel variations. In HDA systems, many precoding algorithms require only second-order

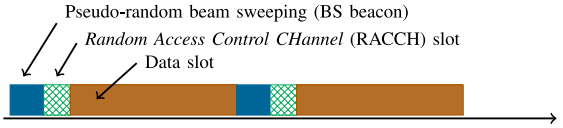

Figure 3.2: Diagram of the frame structure.

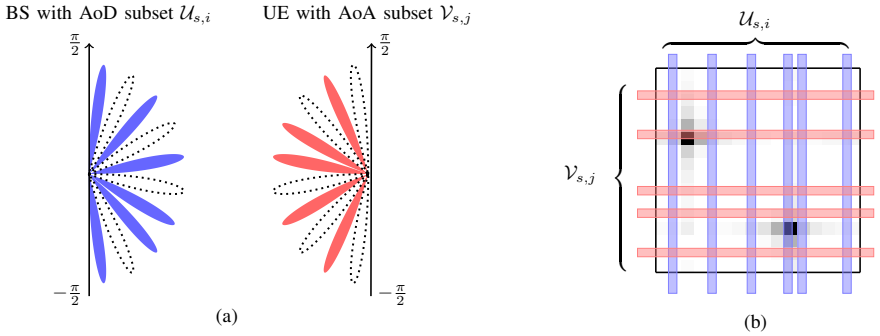

Figure 3.3: Illustrations of the beamforming codebooks for one beacon slot: (a) the beam patterns of the base station and a UE and (b) the channel with two strong paths (indicated by dark spots) probed with the beam patterns.

statistics of the high-dimensional channel (e.g. [19]). An advantage of such precoders is that they relax the requirement on the update rate of the analog network. The beam-domain second-order statistics of the channel of the k-th user $\mathbf{\Gamma}_k$ is an all-zero $N_{\mathsf{UE}} \times N_{\mathsf{BS}}$ matrix with positive elements given by

$$[\mathbf{\Gamma}_k]_{n,d} \propto \sum_{l=1}^{L_k} \mathbb{E}\left[\left|[\check{\mathsf{H}}_{k,s,l}(t,\tau_{k,l})]_{n,d}\right|^2\right], n \in N_{\mathsf{UE}}, d \in N_{\mathsf{BS}}. \tag{3.1}$$

Our proposed algorithm works as follows. It is based on a frame structure as shown in Figure 3.2. The frame consist of three parts: the beacon slot, the random access control channel (RACCH) slot, and the data slot. This frame structure is very similar to the 5G frame structure [20]. The BA process is done during the beacon slot in downlink. As illustrated in our previous work [5], the base station broadcasts a training sequence. The training sequence consists of a series of pseudo-random beam patterns (referred to as the transmit beamforming codebook). The codebook is known by the users. A UE measures the training sequence with its own codebook of receive beam patterns. An illustration of one measurement with selected codewords for the base station and a UE is shown in Fig. 3.3. The beam patterns consist of pseudo-random selections of directions. The number of probed directions is a trade-off between the beamforming gain of the patterns and the necessary training time. The number of measurements 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 report). In the frequency-domain version of the algorithm the base station uses orthogonal OFDM-symbols as training signals. The training signals in the time-domain version are unique pseudo-noise sequences. We assign a unique training sequence to every RF chain / codeword pair of the base station. Over T beacon slots a UE obtains a total number of $N_{\mathsf{RF}}T$ equations, which can be written in the form

$$\mathbf{q}_k = \mathbf{B}_k \cdot \mathrm{vec}(\mathbf{\Gamma}_k) + \zeta(P_{\mathsf{tot}}) \cdot \mathbf{1} + \mathbf{w}_k, \tag{3.2}$$

where $\mathbf{q}_k \in \mathbb{R}^{N_{\mathsf{RF}}T}$ consists of all the $N_{\mathsf{RF}}T$ statistical power measurements, $\mathbf{B}_k \in \mathbb{R}^{N_{\mathsf{RF}}T \times N_{\mathsf{BS}}N_{\mathsf{UE}}}$ is uniquely defined by the pseudo-random beamforming codebook of the base station and the

local beamforming codebook of the k-th UE, $\zeta(P_{\text{tot}})$ denotes a constant whose value is a function of the total radiated power, and $\mathbf{w}_k \in \mathbb{R}^{N_{\text{RF}}T}$ denotes the residual measurement fluctuations. As discussed in [5, 4], with the non-negative constraint of $\mathbf{\Gamma}_k$, a simple least squares

$$\mathbf{\Gamma}_k^{\star} = \underset{\mathbf{\Gamma}_k \in \mathbb{R}_+^{N_{\text{UE}} \times N_{\text{RF}}}}{\arg\min} \|\mathbf{B} \cdot \text{vec}(\mathbf{\Gamma}_k) + \zeta(P_{\text{tot}}) \cdot \mathbf{1} - \mathbf{q}_k\|^2 \tag{3.3}$$

is sufficient to recover the solution $\mathbf{\Gamma}_k^{\star}$. The optimization problem of equation (3.3) is generally called *Non-Negative-Least-Squares* and it has well-investigated numerical solutions. We assume a success of the BA process if the largest component in $\mathbf{\Gamma}_k^{\star}$ coincides with the actual strongest path of the k-th UE. More details can be found in [5, 4].
After the UE has estimated the AoD-AoA pair of its strongest path with the base station, it feeds the AoD information back to the base station. During the RACCH slot, the base station stays in listening mode such that each UE sends a beamformed packet to the base station. This packet contains basic information such as the UE ID and the beam indices of the selected AoD. The base station responds with an acknowledgment data packet in the data subslot of a next frame. From this moment on, the base station 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 mutlipath component with AoA-AoD $(\phi_{k,l}, \theta_{k,l})$ and strong coefficient $\rho_{k,l}$.
We have run numerical simulations of the time-domain version of our algorithm. We compare the performance of both architectures. We considered a system with $N_{\text{BS}} = 64$ antenna elements and $N_{\text{RF}} = 8$ RF chains at the base station. The UE has $N_{\text{UE}} = 8$ antenna elements and one RF chain. The algorithm is independent of the number of users. The antenna array of the FC architecture is one large ULA. The antenna of the OSPS structure consists of N_{RF} disjoint subarrays where each subarray is a ULA. To effectively capture the channel quality before BA, we also define the SNR before beamforming as

$$\text{SNR}_{\text{BBF}} = \frac{P_{\text{tot}} \sum_{l=1}^{L} \gamma_l}{N_0 B}. \tag{3.4}$$

This is the communication data SNR obtained when a single pilot stream ($N_{\text{RF}} = 1$) is transmitted through a single base station antenna element and is received at a single UE antenna element (isotropic transmission) over the whole bandwidth B. P_{tot} is the total radiated power or the effective isotropic radiated power (EIRP) if the base station transmits all its power through one element but scaled with the number of users. The maximum power of the PAs is $P_{\text{max}} = 36\,\text{dBm}$, the backoff factor is $13\,\text{dB}$, and the number of supported users in the system is 8. Assuming $0\,\text{dBi}$ element gain, the total radiated power for the BA process is $P_{\text{tot}} = 36\,\text{dBm} - 13\,\text{dB} + 10\log_{10}(64/8) = 32.03\,\text{dBm}$. The noise power for $B = 800\,\text{MHz}$ is $-83.85\,\text{dBm}$. With a cell size of $300\,\text{m}$ and with a frequency of $40\,\text{GHz}$, the path loss to the cell boundary is $-114.03\,\text{dB}$. This would result in $\text{SNR}_{\text{BBF}} = 1.85\,\text{dB}$. This SNR before beamforming is large, but only considers line-of-sight conditions and neglects the receiver noise figure. We tested the BA algorithm for weaker channel conditions with a $\text{SNR}_{\text{BBF}} = -5\,\text{dB}$ and $\text{SNR}_{\text{BBF}} = -20\,\text{dB}$. Fig. 3.4 (a) and (b) show the detection probability P_D of the BA versus the number of training slots for $-20\,\text{dB}$ and $-5\,\text{dB}$, respectively. The algorithm achieves roughly the same performance for both SNRs. It was parameterized to work in low-SNR regimes. The FC architecture achieves a detection probability of 1 after approximately 40 slots. The OSPS architecture requires roughly 60 to 80 slots. In addition to our algorithm, we also simulated the algorithm proposed in [18]. Fig. 3.4(c) shows the comparison but with slightly different system specifications. The algorithm from the literature does not achieve detection even with 100 training slots. Please see [4] for a detailed comparison.

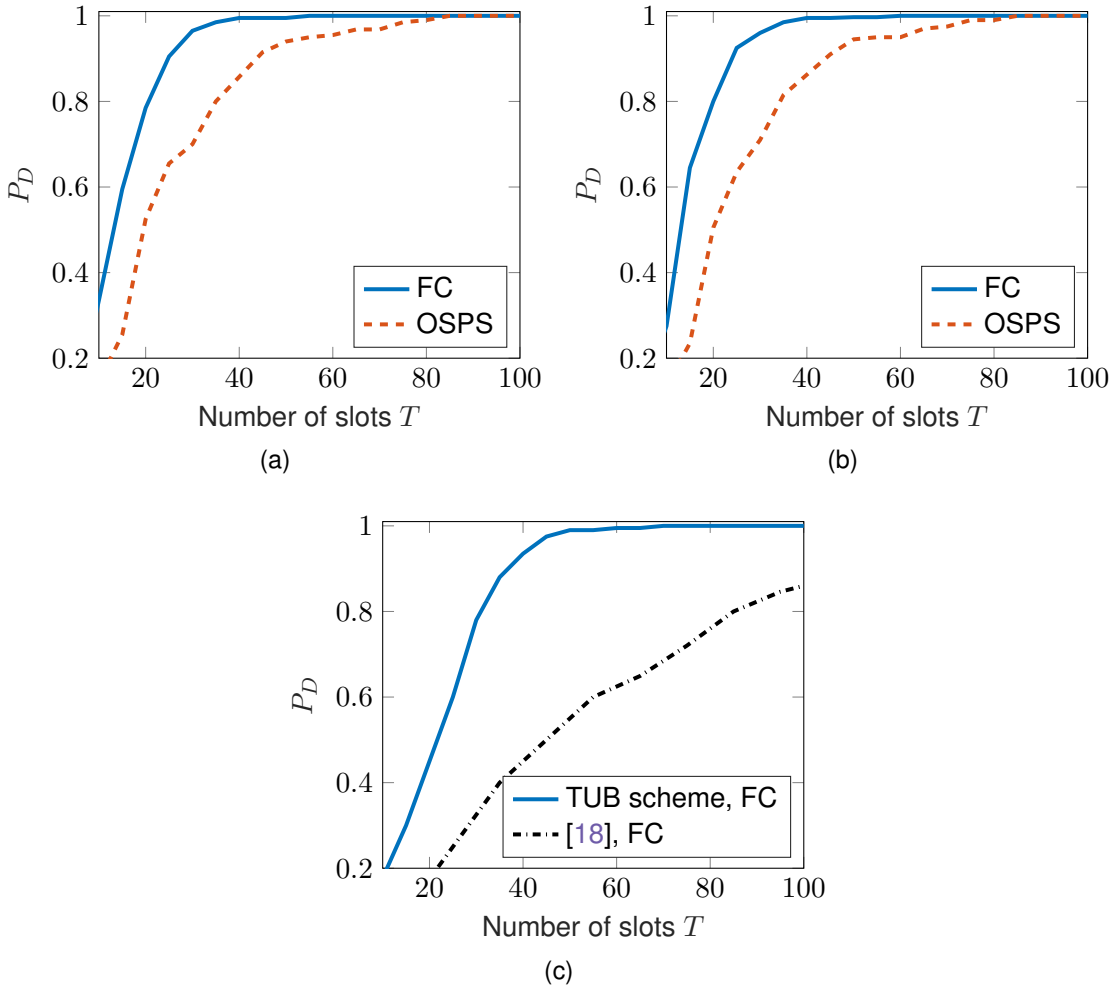

Figure 3.4: Detection probability P_D of different transmitter architectures vs. the training overhead for the initial beam alignment phase with (a) $\text{SNR}_{\text{BBF}} = -20\,\text{dB}$, (b) $\text{SNR}_{\text{BBF}} = -5\,\text{dB}$, and (c) a comparison with the literature.

3.3 Conclusion on the Beam Alignment

The previous two sections introduced algorithms for the BA phase. The exhaustive search can use the RF chains at the base station in parallel and further reduce the required number of slots. With an FC architecture, the training time is $\frac{N_{\text{BS}} N_{\text{UE}}}{N_{\text{RF}}} = 64$. The hierarchical bisection method even without using multiple RF chains only needs 18 slots. Nonetheless, the probability of error would be much larger for the hierarchical algorithm since the beamforming gain of the first level is very low. Additionally, it needs a parallel feedback channel to handle the exchange of the strongest beam per level. The hierarchical search works as described only for a single user. The training time would increase linearly with the number of users. This makes it infeasible for cellular systems. In contrast, the required training time of the proposed advanced algorithm does not change with the number of users. Its assumed frame structure aligns well with the 5G standard. The required training time of 40 slots is lower than the baseline of the exhaustive search, and it scales better with the number of antenna elements (see [5, 4]).Comparing both architectures, the FC structure outperforms the OSPS structure. Nonetheless, the proposed algorithm achieves beam alignment in an acceptable number of training slots.

For the proof-of-concept system, we will implement the exhaustive search and our advanced

algorithm. In this way, we can compare the performance of the advanced algorithm with the baseline of the exhaustive search. If the training time of the exhaustive search is shorter than the channel coherence time, it measures the instantaneous full channel information. In this way, it can be used as a channel sounder.

Chapter 4

Data communication phase

4.1 Precoding algorithm

After the initial acquisition of users for the base station 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 base station to the UE. In the uplink, the UE transmits to the base station which uses MU-MIMO detection. The downlink precoding algorithms presented in the following are reciprocal and can also be used as detection algorithms. As written in the introduction, the system operates in TDD mode. Hence, we will describe and simulate the algorithms for the downlink. Furthermore, we assume that the base station simultaneously schedules $K = N_{\mathsf{RF}}$ users which are selected by a simple directional scheduler [21]. The simple scheduler selects K users which have similar power profiles and whose strongest AoDs are at least $\Delta\theta_{\mathsf{min}}$ 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 allocates equal power across these users. Through the signal processing the base station can apply multi-user interference cancellation [21]. In (4.1), $\mathbf{u}_k$ denotes the normalized transmit beamforming vector for the k-th UE at the base station and $\mathbf{v}_k$ the normalized receive beamforming vector at the k-th UE. The effective radiated power for the k-th user $P_k = \frac{P_{\mathsf{tot}}}{N_{\mathsf{RF}}}$ maintains the total radiated power constraint. The received signal at the k-th UE can then be written as

$$\begin{aligned} y_k(t) =& \mathbf{v}_k^{\mathsf{H}} \sum_{k'=1}^{K} \sqrt{P_{k'}} \mathsf{H}_{k,s}(t,\tau) \circledast (\mathbf{u}_{k'} x_{k'}(t)) + z_k(t) \\ =& \sqrt{P_k} \left(\mathbf{v}_k^{\mathsf{H}} \mathsf{H}_{k,s}(t,\tau) \mathbf{u}_k \right) \circledast x_k(t) + z_k(t) \\ &+ \sum_{k' \neq k} \sqrt{P_{k'}} \left(\mathbf{v}_k^{\mathsf{H}} \mathsf{H}_{k,s}(t,\tau) \mathbf{u}_{k'} \right) \circledast x_{k'}(t) \end{aligned} \tag{4.1}$$

where $f(t) \circledast g(t) = \int f(\tau) g(t-\tau) d\tau$ denotes the convolution operation. As we can see, the first term in (4.1) corresponds to the desired signal at the k-th UE, whereas the last two terms correspond to the noise and interference, respectively. By applying the introduced channel model via substituting (1.1) into (4.1), the received signal reads

$$\begin{aligned} y_k(t) =& \sum_{l=1}^{L_k} \sqrt{P_k} \mathbf{v}_k^{\mathsf{H}} \mathsf{H}_{k,s,l}(t) \mathbf{u}_k x_k(t - \tau_{k,l}) + z_k(t) \\ &+ \sum_{k' \neq k} \sum_{l=1}^{L_{k'}} \sqrt{P_{k'}} \mathbf{v}_k^{\mathsf{H}} \mathsf{H}_{k',s,l}(t) \mathbf{u}_{k'} x_{k'}(t - \tau_{k',l}), \end{aligned} \tag{4.2}$$

where $x_k(t)$ denotes the unit-power transmit signal and $z_k(t) \sim \mathcal{CN}(0, N_0 B)$ denotes the complex additive white Gaussian noise. N_0 is the power spectral density (in W/Hz) and B denotes the effective bandwidth. By treating the multi-user interference as noise at each UE we can use the Shannon–Hartley theorem to derive the maximum spectral efficiency of a user. Hence, the asymptotic spectral efficiency of the k-th UE is given by

$$R_k = \mathbb{E}\left[\log_2\left(1+\frac{P_k|\sum_{l=1}^{L_k}\mathbf{v}_k^{\mathsf{H}}\mathsf{H}_{k,s,l}(t)\mathbf{u}_k|^2}{|\sum\sum_{k'\neq k}^{}\sum_{l=1}^{L_{k'}}\sqrt{P_{k'}}\mathbf{v}_k^{\mathsf{H}}\mathsf{H}_{k',s,l}(t)\mathbf{u}_{k'}|^2+|z_k(t)|^2}\right)\right]. \tag{4.3}$$

The sum spectral efficiency for the base station reads as $R_{\text{sum}} = \sum_{k=1}^{K} R_k$.
Most algorithms proposed in the literature assume full channel state information available at the transmitter (e.g [22, 23, 24]). This assumption does not hold in a realistic system as the training time would decrease the system capacity drastically. In contrast, the following signal processing is based on the outcome of the BA phase described in Section 3.2. More precisely, it is assumed that after a BA procedure the strongest component in $\Gamma_k^\star$ corresponds to the l_k-th multi-path component in $\check{\mathsf{H}}_{k,s}(t,\tau)$ between the base station and the k-th UE. As illustrated in Section 3.2, the estimated beam indices are fed back from the UEs to the base station.
The basic concept of the proposed algorithm is, that the analog networks both at the base station and at the UE point beams towards the strongest path defined by the AoD and AoA. The digital part of the signal processing can maximize the spectral efficiency by applying two schemes described later in this section. To attain the beamforming coefficients the k-th UE decodes its data along the estimated strongest direction, given by

$$\mathbf{v}_k = \mathbf{F}_{\mathsf{UE}}\check{\mathbf{v}}_k, \tag{4.4}$$

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

$$\mathbf{u}_k = \mathbf{F}_{\mathsf{BS}}\check{\mathbf{u}}_k, \tag{4.5}$$

where $\check{\mathbf{u}}_k \in \mathbb{C}^{N_{\mathsf{BS}}}$ is an all-zero vector with a 1 at the component corresponding to strongest AoD direction of the k-th UE.
To formulate the hybrid precoding problem, we re-write everything in a matrix-multiplication format. Let $\mathbf{x}(t) = \operatorname{diag}(\sqrt{P_1}, \sqrt{P_2}, ..., \sqrt{P_K}) \cdot [x_1(t), x_2(t), ..., x_K(t)]^{\mathsf{T}} \in \mathbb{C}^K$ denote the transmit signal vector and $\overline{\mathsf{H}}_s(t,\tau)$ denote the aggregated channel for all the K UEs given by

$$\overline{\mathsf{H}}_s(t,\tau) = \left[\mathsf{H}_{1,s}(t,\tau)^{\mathsf{T}}, \mathsf{H}_{2,s}(t,\tau)^{\mathsf{T}}, \cdots, \mathsf{H}_{K,s}(t,\tau)^{\mathsf{T}}\right]^{\mathsf{T}}, \tag{4.6}$$

where $\mathsf{H}_{k,s}(t,\tau)$, $k \in [K]$, is given in (1.1). We define $\mathbf{V} \in \mathbb{C}^{N_{\mathsf{UE}}K\times K}$ as the receive beamforming matrix given by

$$\begin{aligned}\mathbf{V} &= \operatorname{diag}(\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_K)\\ &= (\mathbf{I}_K \otimes \mathbf{F}_{\mathsf{UE}}) \cdot \operatorname{diag}(\check{\mathbf{v}}_1, \check{\mathbf{v}}_2, ..., \check{\mathbf{v}}_K),\end{aligned} \tag{4.7}$$

where $\mathbf{I}_K$ denotes the $K \times K$ identity matrix and $\otimes$ represents the Kronecker product. Let $\overline{\mathbf{U}} \in \mathbb{C}^{D\times K}$ denote the analog precoding vector support given by

$$\begin{aligned}\overline{\mathbf{U}} &= [\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_K]\\ &= \mathbf{F}_{\mathsf{BS}} \cdot [\check{\mathbf{u}}_1, ..., \check{\mathbf{u}}_K].\end{aligned} \tag{4.8}$$

$\mathbf{A}_B = [\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_K] \in \mathbb{C}^{K\times K}$ denotes the baseband precoding matrix and represents the digital part of the precoding. The complete precoding matrix $\mathbf{U} \in \mathbb{C}^{D\times K}$ at the base station can be written as

$$\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_K] = \overline{\mathbf{U}} \cdot \mathbf{A}_B. \tag{4.9}$$

To meet the total radiated power constraint, the coefficients in (4.9) are normalized as $\|\mathbf{u}_k\| = \|\overline{\mathbf{U}} \cdot \mathbf{a}_k\| = 1$. As a result, the receive signal $\mathbf{y}(t) = [y_1(t), y_2(t), ..., y_K(t)]^{\mathsf{T}} \in \mathbb{C}^K$ reads

$$\begin{aligned}\mathbf{y}(t) &= \mathbf{V}^{\mathsf{H}} \cdot \overline{\mathsf{H}}_s(t,\tau) \circledast (\mathbf{U} \cdot \mathbf{x}(t)) + \mathbf{z}(t) \\ &= \left(\mathbf{V}^{\mathsf{H}} \cdot \overline{\mathsf{H}}_s(t,\tau) \cdot \overline{\mathbf{U}} \cdot \mathbf{A}_B\right) \circledast \mathbf{x}(t) + \mathbf{z}(t) \\ &:= \left(\widetilde{\mathsf{H}}_s(t,\tau) \cdot \mathbf{A}_B\right) \circledast \mathbf{x}(t) + \mathbf{z}(t),\end{aligned} \tag{4.10}$$

where $\mathbf{z}(t) \in \mathbb{C}^K$ denotes the noise, and

$$\widetilde{\mathsf{H}}_s(t,\tau) = \mathbf{V}^{\mathsf{H}} \cdot \overline{\mathsf{H}}_s(t,\tau) \cdot \overline{\mathbf{U}} \tag{4.11}$$

represents the $K \times K$-lower-dimensional effective channel.

Beam steering scheme

The beam steering (BST) scheme is the simplest possible approach for the digital signal processing. The base station ignores the multi-user interference. It directly transmits the data stream for the k-th UE using the k-th RF chain. Therefore, the digital precoding matrix is given by $\mathbf{A}_B^{\mathsf{BST}} = \mathbf{I}_K$. In this case, an additional uplink channel estimation of $\widetilde{\mathsf{H}}_s(t,\tau)$ can be omitted. The eventual $D \times K$ BST precoder in (4.9) reads

$$\mathbf{U}^{\mathsf{BST}} = \overline{\mathbf{U}} \cdot \mathbf{A}_B^{\mathsf{BST}} = \overline{\mathbf{U}}. \tag{4.12}$$

Baseband zeroforcing scheme

In the baseband zeroforcing (BZF) scheme, we consider zeroforcing precoding for potential multi-user interference cancellation. To calculate the zeroforcing precoding matrix $\mathbf{A}_B^{ZF}$ the base station needs to estimate the lower-dimensional effective channel $\widetilde{\mathsf{H}}_s(t,\tau)$. The effective channel can be estimated using channel reciprocity and standard orthogonal uplink pilots at the cost of $(K \ll MN)$ orthogonal pilots. As a result, the baseband precoding matrix $\mathbf{A}_B^{\mathsf{ZF}}$ can be written as

$$\mathbf{A}_B^{\mathsf{ZF}} = \widetilde{\mathsf{H}}_s(t,\tau)^{\mathsf{H}} \cdot \left(\widetilde{\mathsf{H}}_s(t,\tau)\widetilde{\mathsf{H}}_s(t,\tau)^{\mathsf{H}}\right)^{-1} \cdot \Delta^{\mathsf{ZF}}, \tag{4.13}$$

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

$$\mathbf{U}^{\mathsf{ZF}} = \overline{\mathbf{U}}^{\mathsf{ZF}} \cdot \mathbf{A}_B^{\mathsf{ZF}}. \tag{4.14}$$

4.2 Numerical simulations

In this section, we will compare the asymptotic sum spectral efficiency in terms of different transmitter architectures. We again considered a system with $N_{\mathsf{BS}} = 64$ antenna elements and

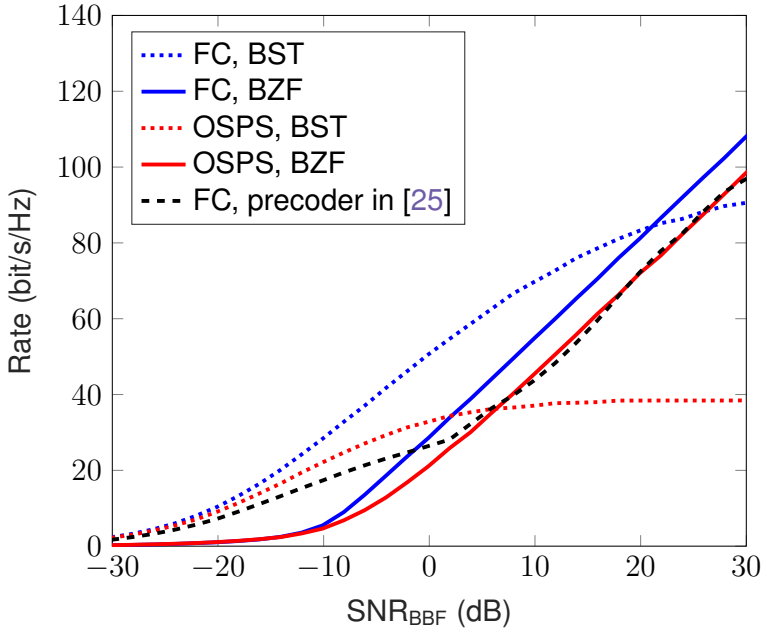

Figure 4.1: The sum spectral efficiency of both architectures vs. increasing $\mathrm{SNR}_{\mathrm{BBF}}$ for the data communication phase with different precoding schemes.

$N_{\mathrm{RF}} = 8$ RF chains at the base station. The $K = 8$ users have $N_{\mathrm{UE}} = 8$ antenna elements and one RF chain each. The antenna array of the FC architecture is one large ULA. The antenna of the OSPS structure consists of N_{RF} disjoint subarrays where each subarray is a ULA. The system is assumed to work at $f_0 = 40$ GHz with a maximum available bandwidth of $B = 0.8$ GHz. We assume that the channel for each UE contains $L_k = 3$ links given by $(\gamma_{k,1} = 1, \eta_{k,1} = 100)$, $(\gamma_{k,2} = 0.6, \eta_{k,2} = 10)$, and $(\gamma_{k,3} = 0.6, \eta_{k,3} = 0)$ as defined in (1.3).
Fig. 4.1 shows the sum spectral efficiency for both architectures and both precoding schemes versus the $\mathrm{SNR}_{\mathrm{BBF}}$. As a comparison, we also simulated a recent precoding algorithm from [25] which is based on full channel reconstruction. In the range of $\mathrm{SNR}_{\mathrm{BBF}} \leq 10$ dB, the simple BST scheme achieves the highest sum spectral efficiency, when $\mathrm{SNR}_{\mathrm{BBF}} \gg 0$ dB, the BZF precoder performs better. The OSPS transmitter using the BST precoding saturates to a lower rate above $\mathrm{SNR}_{\mathrm{BBF}} > -15$ dB compared to the FC structure. Using the BZF precoding the OSPS architecture with $\mathrm{SNR}_{\mathrm{BBF}} > 0$ dB shows a similar performance but with a constant gap compared to the FC structure. The proposed precoders achieve a better performance than that of [25].
In summary, under the given assumptions the OSPS architecture performs only slightly worse compared to the FC architecture. Due to the much lower complexity of the OSPS analog network, this performance loss is acceptable for the SERENA proof-of-concept system. The choice for the precoding scheme strongly depends on the usage scenario, cell size, and user distribution within the cell. Due to the similar structure of the digital precoding a base station system can implement both schemes. The scheduler can choose, depending on the measured SNR, which scheme to use for the data transmission. We will implement both schemes on our digital signal processing platform for the SERENA proof-of-concept system and reevaluate the simulation results with experimentation.

Chapter 5

Conclusion and resulting area capacity

In this deliverable, we gave a general introduction to the signal processing of HDA systems like the SERENA platform. We focused on the parts of the signal processing which are influenced by the HDA structure. These parts are the architecture of the hybrid digital-analog structure, the initial acquisition (also called beam alignment phase), and the MU-MIMO precoding. We listed other specifications of more common parts of the signal processing in the introduction. For these other parts, e.g. the modulation, we will use standard algorithms in the proof-of-concept system.
We introduced the FC and the OSPS HDA architectures. The SERENA platform will be built with an OSPS architecture. In our comparison of the two architectures, we showed that the OSPS architecture has similar or slightly better power efficiency compared to the FC case. The main advantage of an OSPS analog network is the reduced complexity. This is also the main reason to use this structure in the proof-of-concept system. The FC structure would be infeasible with the SERENA technology which is already beyond state of the art. In the BA and data communication chapters, we continued the comparison of both architectures. The OSPS architecture needs a longer time for the BA process and has a lower sum rate during the data communication phase. However, these differences are not as severe as the increase in complexity of the FC structure. In general, the FC architecture is impossible to realize with state-of-the-art technology under economical aspects.
In Chapter 3 we described the initial acquisition phase. In HDA systems, a mobile cell needs a special procedure to acquire users and measure the channel state. We described two basic schemes and proposed our own advanced algorithm. The two basic schemes, exhaustive search and hierarchical bisection search, are not well suited for mobile communications. The exhaustive search requires too many training slots and does not scale well with the number of antennas. The hierarchical bisection algorithm does not work well with many users. We introduced our own advanced algorithm, which is based on the concept of compressed sensing, and simulated its performance. For the proof-of-concept system, we will implement the exhaustive search and the advanced algorithm. The exhaustive search scheme is easy to implement and a good benchmark. It also measures the complete channel. The advanced algorithm will be mainly used to demonstrate the performance of the communication proof-of-concept.
The last part of the HDA signal processing description is the data communication phase. We have proposed and investigated an algorithm for MU-MIMO precoding. It offers two schemes in the digital baseband which enable different performance for different SNR scenarios. We simulated the performance and compared the data rate for both architectures. We will use both schemes for the proof-of-concept system.
From the achieved simulation results we can estimate the downlink area capacity of a cell using the SERENA platform. The system specifications as mentioned in the previous chapters would

be a maximum power of the PAs $P_{\text{max}} = 36\,\text{dBm}$, a backoff factor of $13\,\text{dB}$, 8 users, $N_{\text{BS}} = 64$ base station antenna elements, $N_{\text{RF}} = 8$ base station RF chains, and UEs with $N_{\text{UE}} = 8$ antenna elements and one RF chain. The cell operates at $40\,\text{GHz}$ with a bandwidth of $B = 800\,\text{MHz}$. We assume a cell size of $300\,\text{m}$ and an angular coverage of $90°$ per sector. In such a scenario, the line of sight path loss to the cell boundary is $-114.03\,\text{dB}$. The SNR_{BBF} as introduced in Equation (3.4) would be $1.85\,\text{dB}$. The resulting sum spectral efficiency R from Fig. 4.1 is $34\,\text{bits/s/Hz}$ for the OSPS architecture using the BST precoding scheme. This is of course a theoretical value, which does not take actual modulation or channel coding into consideration. Therefore, we define $\epsilon = 0.6$ as a modulation factor. Please note that the simulation still takes the channel model with multiple scatterers into account and is not pure line of sight. The coverage area in km^2 of the system would be $a = \frac{90°}{360°}\pi(0.3\,\text{km})^2 = 0.071\,\text{km}^2$. The area capacity $C = B\epsilon R/a$ with these assumptions is $C_{\text{SERENA}} = 229.9\,\text{Gbit}\,\text{s}^{-1}\,\text{Hz}^{-1}\,\text{km}^{-2}$.
As a benchmark, we compare our result to an LTE advanced (LTE-A) cell which in a MU-MIMO configuration of 8×8 has a spectral efficiency of $R = 30\,\text{bits/s/Hz}$ [26, p.416] and a bandwidth of $20\,\text{MHz}$. Typical LTE cells are larger. We assume a cell radius of $750\,\text{m}$ and $120°$ per sector. As a result, the area capacity is $C_{\text{LTE}} = 1.02\,\text{Gbit}\,\text{s}^{-1}\,\text{Hz}^{-1}\,\text{km}^{-2}$. With these assumptions, the SERENA platform would improve the area capacity compared to LTE by a factor of over 200. Even if we would increase the assumed cell size for the SERENA system to $750\,\text{m}$ and recalculate the area capacity, we would still achieve $C_{\text{SERENA,2}} = 29.3\,\text{Gbit}\,\text{s}^{-1}\,\text{Hz}^{-1}\,\text{km}^{-2}$. Of course, many of these assumptions are very theoretical. For example, the assumed LTE spectral efficiency is the peak value, not a cell average or a cell boundary value. On the other hand, e.g., the calculations for the SERENA capacity neglected the receiver noise figure, assumed an arbitrary ϵ to account for the modulation and channel coding, and use a simple and sparse channel model. Nonetheless, from the theoretical simulations, we can be confident that SERENA will reach its area capacity goal of an increase by 10 to 100 times given in the DoA.

Chapter 6

List of Abbreviations

Abbreviation	Translation
AoA	Angle of arrival
AoD	Angle of departure
BA	Beam alignment
BST	Beam steering
BZF	Baseband zeroforcing
CSI	Channel state information
DFT	Discrete Fourier transform
EIRP	Effective isotropic radiated power
FC	Fully-connected
HDA	Hybrid digital-analog
IC	Integrated circuit
MU-MIMO	Multiuser multiple-input multiple-output
OFDM	Orthogonal frequency division multiplexing
OSPS	One-stream-per-subarray
PA	Power amplifier
PAPR	Peak-to-average power ratio
QAM	Quadrature amplitude modulation
RACCH	Random access control channel
RF	Radio frequency
RX	Receiver
SC	Single-carrier
SNR	Signal-to-noise ratio
TDD	Time division multiplexing
TX	Transmitter
UE	User equipment
ULA	Uniform linear array

References

[1] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[2] A. F. Molisch, V. V. Ratnam, S. Han, Z. Li, S. L. H. Nguyen, L. Li, and K. Haneda, "Hybrid beamforming for massive MIMO: A survey," *IEEE Communications Magazine*, vol. 55, no. 9, pp. 134–141, 2017.

[3] T. Marzetta, "Massive MIMO: An introduction," *Bell Labs Technical Journal*, vol. 20, pp. 11–22, 2015.

[4] X. Song, S. Haghighatshoar, and G. Caire, "Efficient beam alignment for mmwave single-carrier systems with hybrid MIMO transceivers," *arXiv preprint arXiv:1806.06425*, 2018.

[5] ——, "A scalable and statistically robust beam alignment technique for millimeter-wave systems," *IEEE Transactions on Wireless Communications*, vol. 17, no. 7, pp. 4792–4805, July 2018.

[6] 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.

[7] 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," *arXiv preprint arXiv:1806.01602*, 2018.

[8] 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.

[9] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1164–1179, June 2014.

[10] "Wireless LAN medium access control (MAC) and physical layer (PHY) specifications amendment 3: Enhancements for very high throughput in the 60 GHz band," IEEE Standard 802.11ad, 2014.

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

[12] 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, October 2013.

[13] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831–846, Oct 2014.

[14] K. Sayidmarie and Q. H. Sultan, "Synthesis of wide beam array patterns using quadratic-phase excitations," vol. 3, pp. 127–135, 12 2013.

[15] O. Abari, H. Hassanieh, M. Rodriguez, and D. Katabi, "Millimeter wave communications: From point-to-point links to agile network connections," in *Proceedings of the 15th ACM Workshop on Hot Topics in Networks (HotNets)*. New York, NY, USA: ACM, 2016, pp. 169–175.

[16] T. Kuehne and G. Caire, "An analog module for hybrid massive MIMO testbeds demonstrating beam alignment algorithms," in *WSA 2018; 22nd International ITG Workshop on Smart Antennas*, March 2018, pp. 1–8.

[17] Y. Han and J. Lee, "Two-stage compressed sensing for millimeter wave channel estimation," in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 860–864.

[18] K. Venugopal, A. Alkhateeb, R. W. Heath, and N. G. Prelcic, "Time-domain channel estimation for wideband millimeter wave systems with hybrid architecture," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2017, pp. 6493–6497.

[19] 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.

[20] X. Lin, J. Li, R. Baldemair, T. Cheng, S. Parkvall, D. Larsson, H. Koorapaty, M. Frenne, S. Falahati, A. Grövlen, and K. Werner, "5G new radio: Unveiling the essentials of the next generation wireless access technology," *CoRR*, vol. abs/1806.06898, 2018. [Online]. Available: http://arxiv.org/abs/1806.06898

[21] 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, June 2017.

[22] 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.

[23] 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?" *arXiv preprint arXiv:1806.02857*, 2018.

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

[25] M. R. Castellanos, V. Raghavan, J. H. Ryu, O. H. Koymen, J. Li, D. J. Love, and B. Peleato, "Channel-reconstruction-based hybrid precoding for millimeter-wave multi-user MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 2, pp. 383–398, 2018.

[26] M. Rumney, *LTE and the Evolution to 4G Wireless: Design and Measurement Challenges*. Wiley, 2009.