UNIVERSIDADE FEDERAL DO CEARÁ

CENTRO DE TECNOLOGIA

DEPARTAMENTO DE ENGENHARIA DE TELEINFORMÁTICA

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE TELEINFORMÁTICA

FAZAL-E-ASIM

ARRAY PROCESSING AND PRECODING DESIGN FOR NEXT GENERATION OF WIRELESS COMMUNICATION SYSTEMS

FORTALEZA

2020

FAZAL-E-ASIM

ARRAY PROCESSING AND PRECODING DESIGN FOR NEXT GENERATION OF WIRELESS COMMUNICATION SYSTEMS

Tese apresentada ao Curso de Doutorado em Engenharia de Teleinformática do Programa de Pós-Graduação em Engenharia de Teleinformática do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de doutor em Engenharia de Teleinformática. Área de Concentração: Sinais e Sistemas

Orientador: Prof. Dr. techn. Dr. h. c. Josef A. Nossek
Coorientador: Prof. Dr. Charles Casimiro Cavalcante

FORTALEZA

2020

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

A857a Asim, Fazal-E-.
Array Processing and Precoding Design for Next Generation of Wireless Communication Systems / Fazal-E- Asim. – 2020.
118 f. : il. color.

Tese (doutorado) – Universidade Federal do Ceará, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia de Teleinformática, Fortaleza, 2020.
Orientação: Prof. Dr. techn. Dr. h. c. Josef A. Nossek.
Coorientação: Prof. Dr. Charles Casimiro Cavalcante.

1. Processamento de Sinais. 2. Comunicações Sem-Fio. 3. Estimação de parâmetros do Canal. 4. Ondas milimétricas. 5. Beamforming. I. Título.

CDD 621.38

FAZAL-E-ASIM

ARRAY PROCESSING AND PRECODING DESIGN FOR NEXT GENERATION OF WIRELESS COMMUNICATION SYSTEMS

Thesis defended at the Teleinformatics Engineering Doctorate Program at the Teleinformatics Engineering Post-Graduate Program of the Technology Center at the Federal University of Ceará, as a requirement to obtain the doctor degree in Teleinformatics Engineering. Concentration Area: Signals and Systems.

Approved on: October 07 , 2020

EXAMINING COMMITTEE

Prof. Dr. techn. Dr. h. c. Josef A. Nossek
(Advisor)
Universidade Federal do Ceará, Brazil/
Technical University of Munich, Germany.

Prof. Dr. Charles Casimiro Cavalcante
(Co-Advisor)
Universidade Federal do Ceará, Brazil.

Univ.-Prof. Dr.-Ing. Martin Haardt
Technische Universität Ilmenau, Germany.

Univ.-Prof. Dr. Didier Le Ruyet
Conservatoire National des Arts et Métiers,
France.

Prof. Dr.-Ing. Felix Antreich
Instituto Tecnológico de Aeronáutica, Brazil.

Prof. Dr. André Lima Férrer de Almeida
Universidade Federal do Ceará, Brazil.

To my parents, family and friends.

ACKNOWLEDGEMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 and CNPq (Procs. 309472/2017-2 and 306616/2016-5).

I would like to thank my advisors, Prof. Dr. techn. Dr. h. c. Josef A. Nossek and Prof. Dr. Charles Casimiro Cavalcante for their technical support at every stage of my Ph.D. In short, it would not be possible without their support.

I am also very grateful to Prof. Dr. André Lima Férrer de Almeida and Prof. Dr.-Ing. Felix Antreich for sharing their knowledge and having technical discussions. I am also thankful to Prof. Dr. Tarcisio Ferreira Maciel for his support and discussions during my stay at Wireless Telecommunication Research Group (GTEL).

I would also like to thank all my GTEL colleagues especially, Dr. Igor Moaco Guerreiro, Dr. Lucas Nogueira Ribeiro, Dr. Roberto Pinto Antonioli, Raphael Braga Evangelista, and Bruno Sokal. To the last, but not least, I am thankful to my colleague and friend Alexandre Matos Pessoa for the technical support and fruitful discussions. Finally, my special thanks to my parents and family, who have been with me all the time.

RESUMO

A próxima geração de sistemas de comunicação sem fio promete fornecer uma melhor experiência do usuário em termos de altas taxas de dados, cobertura, confiabilidade e eficiência energética. Uma das soluções candidatas é a combinação viável de ondas milimétricas (mmWave) com a introdução de um grande número de antenas. Por um lado, o uso de mmWave facilitará a implantação de um grande número de antenas, mas do outro lado vai impor um desafio de implementação de hardware com eficiência energética. Portanto, além da eficiência espectral, a eficiência energética será um importante objetivo de design. A introdução de um grande número de antenas na estação base (BS) também complica a estimativa do parâmetro do canal. A estimativa do parâmetro do canal deve ser obtida com alta resolução no equipamento do usuário (UE), pois esses parâmetros precisam ser quantizados antes de ser enviado de volta à BS para pré-codificação. Se os parâmetros do canal não forem estimados com alta precisão, a BS receberá os parâmetros errados com erros de quantização adicionais, resultando na deterioração do desempenho. Esta tese apresenta uma solução de eficiência energética para superar o desafio de implementação de hardware devido à introdução de um grande número de antenas através da introdução da matriz de Butler (BM) no domínio analógico usando a abordagem de deslocamentos de fase analógicos parcialmente conectado (PCAPS). A implantação da BM melhora a implementação do hardware mas torna a estimativa dos parâmetros do canal e a pré-codificação híbrida mais desafiadores. Para atender a esses problemas, o estimador de máxima verossimilhança (ML) é inicialmente derivado para canais com desvanecimento de frequência plano, enquanto uma abordagem de dois estágios é projetada para estimativa de parâmetro unidimensional assumindo canais seletivos em frequência. A primeira etapa é realizada pela proposição da estimação de parâmetros baseada em um algoritmo de grade DFT (PREIDG) para encontrar as estimativas grosseiras, que é usado para inicializar o algoritmo de maximização de expectativa generalizada alternada de espaço (SAGE) para obter estimativas dos parâmetros através da ML. Além disso, o problema é estendido à estimativa bidimensional de parâmetros, que é resolvida pelo algoritmo de dois estágios. No primeiro estágio, um PREIDG modificado é proposto para realizar uma estimativa grosseira que é usada para obter as estimativas de alta resolução dos parâmetros usando o algoritmo SAGE no segundo estágio. O desempenho dos algoritmos de estimativa dos parâmetros são avaliados derivando o limite inferior de Cramér-Rao (CRLB). Finalmente, o algoritmo analógico e de banda base é obtido usando o método do erro quadrático mínimo ponderado médio (WMMSE) de formação de feixe híbrido (HBF).

Palavras-chave: Beamforming. 5G. MIMO massivo. Ondas milimétricas. Estimação de parâmetros do Canal. Pré-codificação analógica e de banda base.

ABSTRACT

The next generation of wireless communication systems promises to provide a better user experience in terms of high data rates, coverage, reliability, and energy efficiency. One of the competing candidates is the viable combination of millimeter-wave (mmWave) with the introduction of a large number of antennas. On one side, the use of mmWave will facilitate the deployment of a large number of antennas but on the other side will impose a challenge of energy-efficient hardware implementation. Therefore, in addition to spectral efficiency, energy efficiency will be an important design goal. Introducing a large number of antennas at the base station (BS) will also complicate the channel parameter estimation. The channel parameter estimation must be obtained with high-resolution at the user equipment (UE), because these parameters need to be quantized before being sent back to the BS for precoding. If the channel parameters are not estimated with high accuracy, the BS will receive the erroneous parameters with additional quantization errors, resulting in deterioration of performance. This thesis presents an energy-efficient solution to overcome the challenge of hardware implementation due to the introduction of a large number of antennas by introducing Butler matrix (BM) in the analog domain using partially connected analog phase shifting (PCAPS) approach. The deployment of BM improves the hardware implementation but makes the channel parameter estimation and hybrid precoding more challenging. To cater to these problems, maximum likelihood (ML) estimator is initially derived for frequency flat fading channels, while a two-stage approach is designed for one-dimensional parameter estimation assuming frequency selective channels. The first stage is accomplished by proposing parameter estimation based on a DFT grid (PREIDG) algorithm to find the coarse estimates, which is used to initialize the space alternating generalized expectation-maximization (SAGE) algorithm to get ML estimates of the parameters. Furthermore, the problem is extended to two-dimensional parameter estimation, which is solved by the two-stage algorithm. In the first stage a modified PREIDG is proposed to perform coarse estimation which is used to obtain the high-resolution estimates of the parameters using the SAGE algorithm in the second stage. The performance of the parameters estimation algorithms is assessed by deriving Cramér-Rao lower bound (CRLB). Finally, the analog and baseband algorithm is obtained using hybrid beamforming (HBF)-weighted minimum mean square error (WMMSE) method.

Keywords: Beamforming. 5G. Massive MIMO. Millimeter Wave. Channel Parameter Estimation. Analog and Baseband Precoding.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

5G	fifth-generation
ABP	auxiliary beam pair
ADC	analog/digital converter
AoA	angle of arrival
AoD	angle of departure
BC	broadcast channel
BCD	block coordinate descent
BM	Butler matrix
BMs	Butler matrices
bpcu	bits per channel use
BS	base station
CAZAC	constant amplitude zero auto correlation
CRLB	Cramér-Rao lower bound
CS	compressive sensing
CSI	channel state information
DFT	discrete Fourier transform
EM	Expectation-Maximization
ESPRIT	estimation of signal parameters via rotational invariant techniques
FCAPS	fully connected analog phase shifting
FFT	fast Fourier transform
FIM	Fisher information matrix
GoB	grid of beams
HBF	hybrid beamforming
IBC	interference broadcast channel
JADE	joint angle and delay estimation
KKT	Karush-Kuhn-Tucker
LOS	line-of-sight
LUT	Look-up table
MF	matched filter
MIMO	multiple-input multiple-output
MISO	multiple-input single-output

ML	maximum likelihood
MMSE	minimum mean square error
mmWave	millimeter-wave
MSE	mean square error
MU-MIMO	multi-user MIMO
MU-MISO	multi-user MISO
MUSIC	multiple signal classification
NLOS	non-line-of-sight
OMP	orthogonal matching pursuit
PA	power amplifier
PCAPS	partially connected analog phase shifting
PCBM	partially connected Butler matrix
PCBMS	partially connected Butler matrices
pdf	probability density function
PREIDG	parameter estimation based on a DFT grid
PS	phase shifter
PSs	phase shifters
QPSK	quadrature phase shift-keying
RC	raised cosine
RF	radio-frequency
RMSE	root mean square error
SAGE	space alternating generalized expectation-maximization
SFG	signal-flow graph
SINR	signal-to-interference and noise ratio
SNR	signal-to-noise ratio
SU-MISO	single-user MISO
UE	user equipment
ULA	uniform linear array
URA	uniform rectangular array
WMMSE	weighted minimum mean square error
ZF	zero forcing

NOTATION

$\jmath$	imaginary unit $\sqrt{-1}$
a	scalar
$\mathbf{a}$	vector
$\mathbf{A}$	matrix
$\mathbb{R}$	the real numbers field
$\mathbb{C}$	the complex numbers field
$(\cdot)^*$	complex conjugate
$(\cdot)^{\mathsf{T}}$	transpose
$(\cdot)^{\mathsf{H}}$	conjugate transpose (Hermitian)
$(\cdot)^{-1}$	inverse operator
$(\cdot)^{\dagger}$	Moore-Penrose pseudo-inverse
$\|\cdot\|_2$	ℓ_2 norm
$\|\cdot\|_{\mathrm{F}}$	Frobenius norm
$\mathbb{E}[\cdot]$	statistical expectation
$\|\cdot\|$	absolute value (complex magnitude)
$\lfloor\cdot\rfloor$	floor operation
$\otimes$	Kronecker product
$\circ$	outer product
$\odot$	elementwise (Hadamard) product
$\diamond$	Khatri-Rao product
$O(\cdot)$	big-O notation
$\mathbf{I}_N$	N-dimensional identity matrix
$\mathbf{0}_{M\times N}$	$(M \times N)$-dimensional null matrix
$\mathbf{1}_{M\times N}$	$(M \times N)$-dimensional all ones matrix
$\mathrm{tr}\{\cdot\}$	matrix trace
$\det(\cdot)$	matrix determinant
$\mathrm{diag}\{\cdot\}$	transforms an input vector into a diagonal matrix

$\mathrm{vec}(\cdot)$	vectorizes the input array
$\mathrm{unvec}(\cdot)$	reshapes a column vector into a matrix
$\mathrm{U}(a, b)$	uniform distribution from a to b
$\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	complex Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$
$[\mathbf{A}]_{p,q}$	entry of $\mathbf{A}$ at the pth row and qth column
$\max(\mathbf{a})$	gives the index of the maximum value in vector $\mathbf{a}$
$\mod(\cdot)$	modulus

CONTENTS

1 INTRODUCTION

1.1 Motivation

The next generation of wireless communication systems envisioned to support very high data rates for the UEs. To turn this ambitious goal into reality, the use of high bandwidth is essential, which is only possible by moving to the mmWave bands [1, 2]. The use of such higher frequencies will facilitate the deployment of massive multiple-input multiple-output (MIMO) at the BS to achieve high beamforming gains. The available large bandwidth combined with a large number of antennas directly imposes challenges for hardware implementation of such a system [3]. Consequently, in addition to spectral efficiency, energy efficiency becomes an important design goal.

To combat this issue, multiple solutions are proposed to ensure the practicality of the large antenna systems, as the use of one radio-frequency (RF) chain per antenna seems not to be a feasible solution with high resolution analog/digital converter (ADC)s. One of the proposed solutions is the use of low resolution ADC [4, 5, 6]. Another solution is to divide the full digital beamforming into two parts, i.e., analog beamforming and digital baseband precoding [7, 8, 9, 10]. There are two main approaches of hybrid beamforming, namely fully connected analog phase shifting (FCAPS) and partially connected analog phase shifting (PCAPS) network, both of which have pros and cons. In both the strategies dividers are used which divide the signals form the RF-chains to the multitude of PSs. Keep in mind, that the dividers are theoretical lossless, for instance, Wilkinson splitter. The combiners are used only in FCAPS network, which combines the signals from many PSs before the power amplifier (PA) of each antenna which introduces losses. These losses are channel-dependent and depend on the mutual correlation of the precoded signals, which can have a maximum loss of $3\,\mathrm{dB}$, if the two signals are uncorrelated. Furthermore, the total number of PSs required in FCAPS approach is more as compared to PCAPS network, while the using FCAPS, we can form beams by using all the antennas simultaneously.

To cater to this energy losses and to ease the hardware complexity of so many PSs, we introduced BM, which uses fixed PSs and can be constructed with a theoretical lossless 90° hybrid coupler [11, 12, 13]. A BM with low insertion loss is shown in [14]. A BM is the analog implementation of the DFT matrix, which one side provides the energy-efficient hardware implementation in the analog domain but on the other side makes the channel parameter estimation and precoding more challenging.

The mmWave channel estimation becomes more challenging by the introduction of a massive number of antennas at the BS due to its high dimensionality, and it can be even more complicated by introducing the frequency selectivity [15, 16, 17]. Furthermore, the mmWave range is characterized by the sparsity in the angular domain, which facilitates the design of low-complexity compressive sensing (CS) based parameter estimation techniques. Recently, many techniques have been studied for channel parameter estimation by exploiting the sparsity of the channel [18, 19, 20, 21]. An adaptive codebook for channel parameter estimation via CS [22], while an orthogonal matching pursuit (OMP) approach is used in [23] to estimate the channel parameters. By introducing quantized PSs in [24], channel parameter estimation is achieved using frequency selective channels. A two-stage algorithm is introduced in [25] for position and orientation estimation. The authors in [26, 27] proposed a so called beam pairing approach, where many beam pairs are formed, to calculate set of many ratios to estimate the corresponding AoD and angle of arrival (AoA), respectively. The ABP method outperforms grid of beams (GoB) method [28].

The authors in [29] proposed a super-resolution algorithm for flat-fading channels, and further showed that the ABP method outperforms the standard CS techniques based on adaptive codebook and OMP. The work in [30] proposed two high-resolution algorithms based on estimation of signal parameters via rotational invariant techniques (ESPRIT), named as two-dimensional ESPRIT and the minimum searching for frequency selective channels. However, the algorithms are unable to show good performance in the low signal-to-noise ratio (SNR) regime which is important for mmWave fifth-generation (5G) systems. Indeed, ESPRIT is more sensitive to sensor location errors and mutual coupling because of the assumption of translational invariance of subarray responses [31, 32]. However, the SAGE algorithm does not impose such kind of restriction on the subarray responses and outperforms multiple signal classification (MUSIC), ESPRIT and joint angle and delay estimation (JADE) algorithms [33, 34, 35, 36, 37]. Most of the aforementioned work are restricted to uniform linear array (ULA) and frequency flat fading channels.

After the BS gets the estimates of the channel parameters, the next task is the optimal precoding to maximize the spectral efficiency of the users and minimize the inter-user interference. As mentioned earlier, one of the approaches is hybrid beamforming where analog beamforming can be implemented using PSs [38], switches[39] and lenses [40]. The authors in [41] used an alternating and manifold optimization based hybrid beamforming algorithm which put constraint of constant modulus on the design of analog phase shifter (PS). Similarly, the work in [42]

assumes variable PSs in the design of analog and baseband precoding. The authors in [43] analyses the energy efficiency of FCAPS and PCAPS assuming realistic RF modeling. The work in [44], designed hybrid beamforming and combining strategies, wherein the first approach, the combiners are designed separately, while the analog and baseband precoders are jointly optimized by minimizing the system sum-MSE. In the latter, an iterative WMMSE algorithm is used to jointly optimized both the analog and digital precoders/combiners.

1.2 Contributions and Thesis Organizations

This thesis focus on two very important problems, one is channel parameter estimation and the other is analog RF and digital baseband precoding. As discussed earlier, to have an energy-efficient hardware implementation, a BM is introduced to excite an ULA, which is the analog implementation of a DFT matrix. A ML estimator is derived to achieve efficient channel parameter estimates for frequency-flat fading channels. The same setup is extended for frequency-selective channels, which makes it a multidimensional, non-linear optimization problem. To achieve an ML estimates, a two-step estimation algorithm is proposed, where in the first stage, the coarse estimation is done via PREIDG algorithm followed by SAGE.

In the next part of the thesis, the same problem is transformed into a two-dimensional channel parameter estimation, where the URA is feed with multiple BMs, where the two-dimensional beams are formed as the result of Kronecker products of DFT vectors. A two-step method is proposed to achieve high-resolution estimates of the channel parameters. In the first stage, a coarse estimation is done via a modified PREIDG algorithm, which is used to initialize the SAGE algorithm to obtain high-resolution of the channel parameters.

In the final part of the thesis, analog and baseband precoding is proposed in a two-step approach, where in the first step a sub-optimal algorithm is proposed to design the analog precoding, which is used in the second step to find the baseband precoder via HBF-WMMSE method, respectively. The outlines of the chapters are given as follows:

- Chapter 2 addresses the one-dimensional parameter estimation both for frequency flat and selective channels. A BM is introduced in the analog domain to excite the ULA. In the first part of the chapter ML estimation criteria are derived, while in the final part of the chapter, a two-stage algorithm is proposed to achieve ML estimates. Finally, CRLB is derived to assess the performance of ML estimates.
- Chapter 3 addresses the extension of the same problem to two-dimensional pa-

rameter estimation by exciting URA with multiple BMs. A two-stage algorithm is proposed to obtain high-resolution channel parameter estimates. In the first stage, modified PREIDG is used to get the coarse estimates of the channel parameters, which is used to initialize the SAGE algorithm to achieve the high-resolution of the parameter estimates. Moreover, CRLB is derived to assess the performance of the SAGE algorithm.

- Chapter 4 addresses the analog and baseband precoding of the same system. In the first step a sub-optimal algorithm is proposed for the design of analog precoding, which is taken into account for the design of baseband precoder using HBF-WMMSE algorithm in the second step.

The thesis chapters are meant to be self-contained and the reader can read them independently without loss of significant information. We explain the derivations of ML estimator, SAGE and CRLB in Appendix A. The derivations of two-dimensional channel parameter estimation using SAGE and CRLB are explained in Appendix B. Finally, the derivations regarding digital baseband precoding is given in Appendix C.

1.2.1 Publications

(J1) **Asim, F.**; Antreich, F.; Cavalcante, C. C.; de Almeida, A. L. F.; Nossek, J. A. Two-dimensional channel parameter estimation for millimeter-wave systems using Butler matrices. ***submitted to*****, IEEE Transactions on Wireless Communications** , June, 2020;

(J2) **Asim, F.**; Antreich, F.; Cavalcante, C. C.; de Almeida, A. L. F.; Nossek, J. A. Channel parameter estimation for millimeter-wave cellular systems with hybrid beamforming. **Signal Processing**, July, 2020;

(J3) **Asim, F.**; de Almeida, A. L. F.; Haardt, M.; Cavalcante, C. C.; Nossek, J. A. Rank-one detector for Kronecker-structured constant modulus constellations. **IEEE Signal Processing Letters**, July, 2020;

(C1) **Asim, F.**; Nossek, J. A.; Antreich, F.; Cavalcante, C. C.; de Almeida, A. L. F. Maximum likelihood channel estimation for millimeter-wave MIMO systems with hybrid beamforming. **IEEE International Workshop on Smart Antennas (WSA)**, April, 2019;

(C2) **Asim, F.**; de Almeida, A. L. F.; Haardt, M.; Cavalcante, C. C.; Nossek, J. A.

Multi-linear encoding and decoding for MIMO systems. **IEEE 11th Sensor Array and Multichannel Signal Processing Workshop (SAM)**, May, 2020;

2 ONE-DIMENSIONAL CHANNEL PARAMETER ESTIMATION

This chapter introduces channel parameter estimation (angles of departure, complex-path gains, and time-delays) using ML estimation for a single path scenario as well as for multipath scenarios based on energy efficient, analog implementation. For the multipath case, SAGE is used to get the ML estimates. In this chapter, we derive CRLB to assess the performance of ML estimates. We also compare the ML estimates with the state of the art ABP method. The proposed algorithm outperforms the ABP method.

2.1 Overview

To make the dream of high data rates as promised in 5G true, the use of mmWave with MIMO system is essential. The large bandwidth available in the mmWave range allows us to work with a large number of antennas but impose a challenge of energy-efficient hardware implementation [3]. One hope of reducing implementation complexity is the introduction of hybrid beamforming, which is dividing the full digital beamforming into analog beamforming implemented at RF and digital baseband precoding. In hybrid beamforming, a small number of RF-chains are used to control a large number of antennas, ending with two types of hybrid beamforming structures, one is using FCAPS network and the other is using PCAPS network.

In the FCAPS network, all the N transmit antennas are controlled by all the $N_{RF} = N_s$ RF-chains as shown in Figure 2.1, where N_s is the number of data streams. In the FCAPS network,

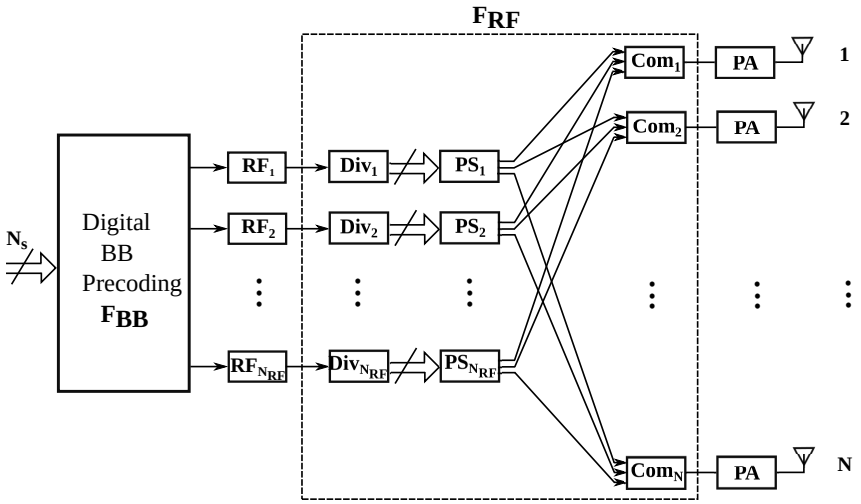


Figure 2.1 – Fully connected analog phase shifting network.

analog signal dividers and combiners are implemented to transmit the part of each data stream through different antennas using $N_{RF}N$ phase shifters (PSs). All the passive elements introduce

losses, especially at mmWave frequencies [45], while power dividers are theoretically lossless e.g. Wilkinson divider, but the combiners are certainly not. The combining losses depend on the mutual correlation of the signals to be combined, for instance, a power loss of $q \times 3$ dB will be witnessed by combining $N_s = N_{RF} = 2^q$ uncorrelated data streams. In addition to energy inefficient implementation, using large number of antennas simultaneously can have problem with far field assumption.

To avoid these significant losses in the analog domain, the PCAPS network is introduced where a combination of N_s signals is avoided in the RF-domain. In PCAPS, the total number of N transmit antennas is divided into multiple sub-arrays such that $N = N_{RF}M$ as shown in Figure 2.2. In this approach, each sub-array gets its signal from one RF-chain, therefore, no combiner is required before transmitting the signal. With this architecture, not only the combining losses will be avoided but the number adaptive PSs will be reduced from NN_{RF} to MN_{RF}.

Moreover, adaptive PSs will need switches to adapt the actual phase shift, therefore both the phase shifter itself, for instance, if it is implemented passively by using a piece of transmission line and the switches will introduce losses, especially at mmWave frequencies.

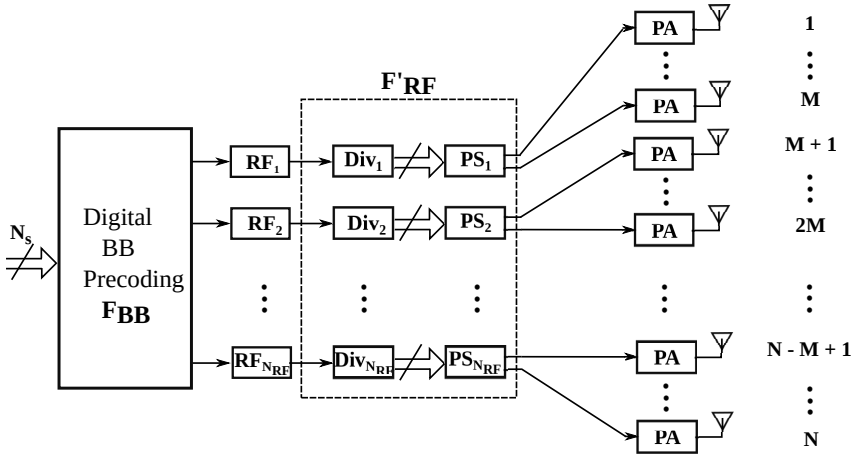


Figure 2.2 – Partially connected analog phase shifting network.

Furthermore, to avoid the implementation of adaptive PSs, we introduce the BM, which is the analog implementation of the DFT matrix with fixed PSs. Using a BM in PCAPS network is shown in Figure 2.3. The M beams of DFT matrix implemented using $\frac{M}{2}\log_2 M$ 90° hybrids with a number of fixed PSs. The 90° hybrid is a theoretically lossless 4-port which can

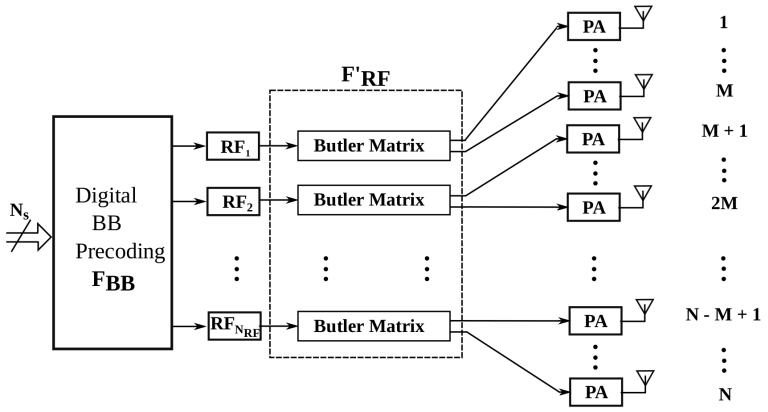


Figure 2.3 – Partially connected Butler matrix approach.

be described by the scattering matrix

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & -j & -1 & 0 \\ -j & 0 & 0 & -1 \\ -1 & 0 & 0 & -j \\ 0 & -1 & -j & 0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} \tag{2.1}$$

and is shown in Figure 2.4. Properly terminating all the ports with no incoming waves at ports 2 and 3 $(a_2 = 0, a_3 = 0)$, which results with no reflected waves at ports $(b_1 = 0, b_4 = 0)$, therefore (2.1) reduces to

$$\begin{bmatrix} b_2 \\ b_3 \end{bmatrix} = \frac{-1}{\sqrt{2}} \begin{bmatrix} j & 1 \\ 1 & j \end{bmatrix} \begin{bmatrix} a_1 \\ a_4 \end{bmatrix}. \tag{2.2}$$

Now, take an example with an input power P_{in} to the $90°$ hybrid at input ports 1 and 4 as P_1 and

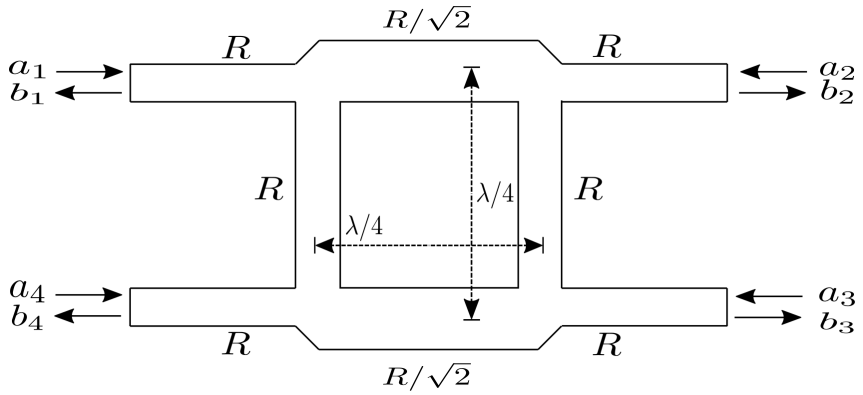


Figure 2.4 – $90°$ hybrid coupler.

P_4 while output ports are 2 and 3, respectively. The $90°$ hybrid will divide the incoming power to the output ports. The input power P_{in} can be described as

$$P_{in} = P_1 + P_4 = \mathbb{E}\left[|a_1|^2\right] + \mathbb{E}\left[|a_4|^2\right] \tag{2.3}$$

and the output power at port 2, P_{out_2} as

$$P_{out_2} = \mathbb{E}\left[|b_2|^2\right] = \frac{1}{2}\left(\mathbb{E}\left[|a_1|^2\right] + \mathbb{E}\left[|a_4|^2\right] - 2\mathrm{Im}\{\rho\}\sqrt{\mathbb{E}\left[|a_1|^2\right]\mathbb{E}\left[|a_4|^2\right]}\right), \tag{2.4}$$

where ρ is known as correlation coefficient

$$\rho = \frac{\mathbb{E}\left[a_1 a_4^*\right]}{\sqrt{\mathbb{E}\left[|a_1|^2\right]\mathbb{E}\left[|a_4|^2\right]}}. \tag{2.5}$$

If the signals at ports 1 and 4 are uncorrelated $\rho = 0$, this results in $P_{out_2} = \frac{1}{2}P_{in}$, which shows that half of the power is lost. However, implementing the BM architecture always cares of using the two output signals and hence no power is lost. An example BM structure with $M = 8$ is shown Figure2.5.

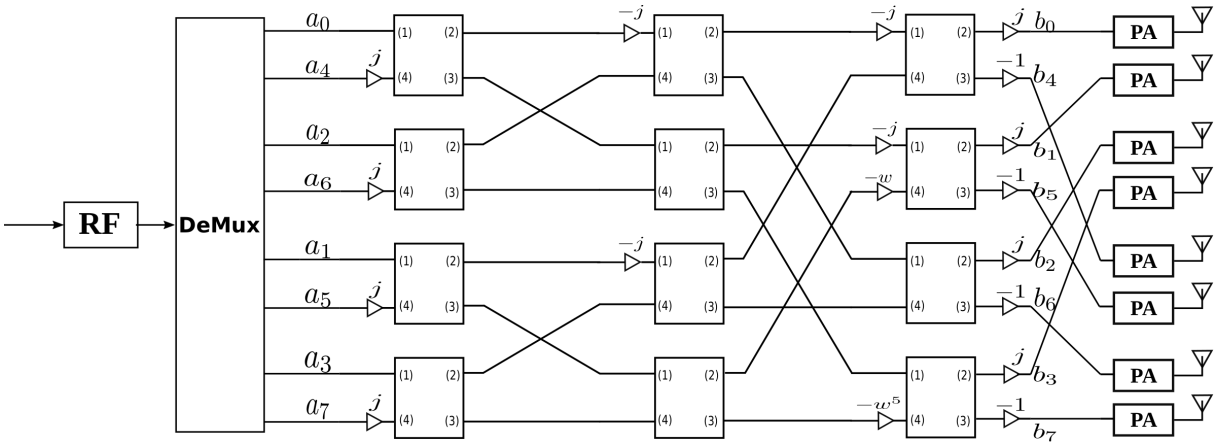


Figure 2.5 – A $M \times M$ BM with $\frac{M}{2}\log_2 M$, 90° hybrids and fixed PSs only, where $w = \exp(-j\frac{2\pi}{M})$ for $M = 8$.

2.2 Contributions

This chapter is organized as follows

- **System Model:**

  We start this chapter with the system model based on the hardware constraints assuming frequency flat and selective channels.

- **ML estimation:**

  In this section, we derive an ML estimator for the estimation of the frequency flat fading channel parameters. Further, we derive the CRLB to assess the performance of our ML estimates.

- **Coarse estimation based on PREIDG algorithm:**

  In this section, the coarse estimation is performed based on the DFT grid for the initial estimation of the frequency selective channel parameters.

- **SAGE:**

  We derive an approximate ML estimator for the non-linear problem to get the refined estimate of the channel parameters using the coarse estimates as an initial guess to reduce the number of iterations for the convergence of SAGE.

- **CRLB:**

  To assess the performance of our so-called two-stage channel parameter estimation algorithm, we derive CRLB bounds.

- **Simulation results:**

  Finally, we will show the performance of an ML estimator for frequency flat and selective fading channels in practical mobile communication scenarios. We will also discuss the simulation results of the two-stage estimation algorithm.

  The author's research contributions include:

  1. Design of the DFT beamformer and its implementation using BM which is an energy-efficient having less implementation complexity in the analog domain.
  2. Derivation of the ML estimator for frequency flat fading channel. Furthermore, the derivation of CRLB to assess the performance of ML estimator.
  3. A two-stage estimation algorithm for frequency selective channel is proposed, where the first stage is the above mentioned coarse estimation based on PREIDG algorithm.
  4. In the final stage, the SAGE algorithm is applied for our configuration which is then initialized by the coarse estimate of parameters to reduce the number of iterations for the convergence and to improve the estimation accuracy.
  5. Derivation of the CRLB for the assessment of the two-stage algorithm.

2.3 ML Estimation for Frequency Flat Fading Channel

In this section, the ML estimator for the frequency flat fading channel is proposed. To assess the performance, we further derive the CRLB.

2.3.1 System Model

We consider a multiuser downlink scenario where the user equipment UE has a single antenna and the base station BS is equipped with a total number of N antennas. The transmitter antennas N are grouped in a sub-array structure as shown in Figure 2.3 as $N = N_{RF}M$. The received signal of one user over a frequency flat fading channel excited by one sub-array is given as

$$\mathbf{y}_k^{\mathrm{T}}[n] = \sqrt{P_T}\alpha\ \mathbf{a}^{\mathrm{H}}(\mu)\mathbf{w}(\Phi_k)\mathbf{c}_k^{\mathrm{T}}[n] + \mathbf{n}_k^{\mathrm{T}} \in \mathbb{C}^{1\times L}. \tag{2.6}$$

For notational convenience, we dropped the user index. P_T is the transmit power of the BS, α is the complex path gain and μ is the spatial frequency given as $\mu = 2\pi\frac{d}{\lambda}\sin\theta$, where d is the distance between antenna elements, λ is the wavelength of the carrier frequency and θ is the AoD. The channel steering vector $\mathbf{a}^*(\mu)$ for the ULA of one sub-array is given as

$$\mathbf{a}(\mu) = [1, e^{-j\mu}, \ldots, e^{-j(M-1)\mu}]^{\mathrm{T}} \in \mathbb{C}^{M\times 1}. \tag{2.7}$$

$\Phi_k = \frac{2\pi}{M}k, k = 0, \ldots, M-1$ is the beamforming angle of the k-th beamforming vector $\mathbf{w}(\Phi_k)$ and is represented as

$$\mathbf{w}(\Phi_k) = \frac{1}{\sqrt{M}}\left[1, e^{-j\Phi_k}, , \ldots, e^{-j(M-1)\Phi_k}\right]^{\mathrm{T}} \in \mathbb{C}^{M\times 1}. \tag{2.8}$$

The constant amplitude zero auto correlation (CAZAC) sequence for each beamforming vector is represented as $\mathbf{c}_k$, where the symbols of each CAZAC sequence are constructed as

$$c(n) = e^{\left(j\frac{2\pi}{\sqrt{L}}\left(\mathrm{mod}\left\{n,\sqrt{L}\right\}+1\right)\left(\left\lfloor\frac{n}{\sqrt{L}}\right\rfloor+1\right)+j\frac{\pi}{4}\right)}, \tag{2.9}$$

where $n \in \{0, 1, \ldots, L-1\}$ and $\mathbf{c}_k(n) = \mathbf{c}(n-k)$. If we consider the length of the sequence $L = 16$, then it creates quadrature phase shift-keying (QPSK) symbols $c(n) \in \left\{\frac{1}{\sqrt{2}}(\pm 1 \pm j)\right\}$, which helps to operate the PA near the saturation region. The CAZAC sequence corresponding to the zeroth beamforming vector is

$$\mathbf{c}_0 = [c(0), c(1), \ldots, c(L-1)]^{\mathrm{T}} \in \mathbb{C}^{L\times 1}, \tag{2.10}$$

and the rest of the CAZAC sequences are just shifted wrap around versions of $\mathbf{c}_0$, while each wraparound is assigned to a specific beamforming vector as $\mathbf{c}_k$, where $k = 0, \ldots, M-1$. The

received matrix $\mathbf{Y}$ can be written by collecting all the received vectors $\mathbf{y}_k$ as,

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_0^{\mathrm{T}} \\ \mathbf{y}_1^{\mathrm{T}} \\ \vdots \\ \mathbf{y}_{M-1}^{\mathrm{T}} \end{bmatrix} = \sqrt{P_T}\alpha\mathbf{A}(\mu) \underbrace{\begin{bmatrix} \mathbf{c}_0^{\mathrm{T}} \\ \mathbf{c}_1^{\mathrm{T}} \\ \vdots \\ \mathbf{c}_{M-1}^{\mathrm{T}} \end{bmatrix}}_{=\mathbf{C}} + \underbrace{\begin{bmatrix} \mathbf{n}_0^{\mathrm{T}} \\ \mathbf{n}_1^{\mathrm{T}} \\ \vdots \\ \mathbf{n}_{M-1}^{\mathrm{T}} \end{bmatrix}}_{=\mathbf{N}} \in \mathbb{C}^{M\times L}$$

where $\mathbf{A}(\mu) = \mathrm{diag}\{\mathbf{a}^{\mathrm{H}}(\mu)\mathbf{w}(\Phi_k)\}_{k=0}^{M-1}$.

2.3.2 ML Estimation

We now probe the frequency flat fading channel by employing all M beamforming vectors, one at a time with the strict correspondences to the CAZAC sequence preferably $L = M$. We treat the observed data matrix $\mathbf{Y}$ as a random variable having a Gaussian probability density function (pdf) parametrized by the parameter $\boldsymbol{\eta}$

$$\boldsymbol{\eta} = [\sigma_n^2, \sqrt{P_T}\alpha, \mu]^{\mathrm{T}} \tag{2.11}$$

with the likelihood function

$$\mathrm{L}(\mathbf{Y};\boldsymbol{\eta}) = \frac{1}{\pi^{M^2}\det\mathbf{R}} \exp\left(-\mathrm{vec}\left\{\mathbf{Y} - \sqrt{P_T}\alpha\mathbf{A}(\mu)\mathbf{C}\right\}^{\mathrm{H}} \mathbf{R}^{-1} \mathrm{vec}\left\{\mathbf{Y} - \sqrt{P_T}\alpha\mathbf{A}(\mu)\mathbf{C}\right\}\right), \tag{2.12}$$

assuming temporally and spatially uncorrelated entries in $\mathbf{N}$,

$$\mathbf{R} = \mathrm{E}\left[\mathrm{vec}\{\mathbf{N}\}\ \mathrm{vec}\{\mathbf{N}\}^{\mathrm{H}}\right] = \sigma_n^2\mathbf{I}_{M^2}. \tag{2.13}$$

Taking the natural logarithm on both sides of (2.12), leads to the log-likelihood function

$$\ell(\mathbf{Y},\boldsymbol{\eta}) = \log_e(\mathrm{L}(\mathbf{Y};\boldsymbol{\eta})) = -M^2\log_e(\pi\sigma_n^2) - \frac{1}{\sigma_n^2}\mathrm{tr}\left\{(\mathbf{Y} - \sqrt{P_T}\,\alpha\ \mathbf{A}(\mu)\mathbf{C}\right)\left(\mathbf{Y} - \sqrt{P_T}\,\alpha\ \mathbf{A}(\mu)\mathbf{C}\right)^{\mathrm{H}}\Bigg\} \tag{2.14}$$

and to the ML estimator

$$\hat{\boldsymbol{\eta}} = \arg\max_{\boldsymbol{\eta}}\left\{\ell(\mathbf{Y},\boldsymbol{\eta})\right\}. \tag{2.15}$$

Now, differentiating (2.14) with respect to σ_n^2 and put it equal to zero results in

$$\hat{\sigma}_n^2 = \frac{1}{M^2}\mathrm{tr}\left\{\left(\mathbf{Y} - \sqrt{P_T}\,\alpha\ \mathbf{A}(\mu)\mathbf{C}\right)\left(\mathbf{Y} - \sqrt{P_T}\,\alpha\ \mathbf{A}(\mu)\mathbf{C}\right)^{\mathrm{H}}\right\}. \tag{2.16}$$

Similarly, we get the expression for the complex path gain $\sqrt{\hat{P_T}}\alpha$ as

$$\sqrt{\hat{P_T}}\alpha = \frac{\mathrm{tr}\{\mathbf{Y}\mathbf{C}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}(\mu)\}}{M\mathrm{tr}\{\mathbf{A}(\mu)\mathbf{A}^{\mathrm{H}}(\mu)\}} \tag{2.17}$$

and for the spatial frequency $\hat{\mu}$

$$\hat{\mu} = \arg\min_{\mu} \mathrm{tr}\left\{\mathrm{diag}\left(\hat{\mathbf{A}} - \mathbf{A}(\mu)\right)\mathrm{diag}\left(\hat{\mathbf{A}} - \mathbf{A}(\mu)\right)^{*}\right\}. \tag{2.18}$$

where $\hat{\mathbf{A}}$ is given as

$$\hat{\mathbf{A}} = \frac{1}{M}\mathbf{Y}\mathbf{C}^{\mathrm{H}}. \tag{2.19}$$

The proof of the derived equations can be found in Appendix A.1.

2.3.3 CRLB

In this section, we derive the CRLB to assess the performance of our propose ML estimator. The parameter vector $\boldsymbol{\eta}$ (2.11) can be re-defined as

$$\boldsymbol{\eta} = \left[\sqrt{P_T}\mathrm{Re}\{\alpha\}, \sqrt{P_T}\mathrm{Im}\{\alpha\}, \mu\right]^{\mathrm{T}}, \tag{2.20}$$

where the complex path gain α is separated into its real and imaginary parts. The FIM for the complex data is given as [46]

$$[\mathbf{F}(\boldsymbol{\eta})]_{ij} = \frac{2}{\sigma_n^2}\ \mathrm{Re}\ \left(\mathrm{tr}\left\{\frac{\partial\mathbf{S}^{\mathrm{H}}(\boldsymbol{\eta})}{\partial\eta_i}\ \frac{\partial\mathbf{S}(\boldsymbol{\eta})}{\partial\eta_j}\right\}\right) \tag{2.21}$$

where $\mathbf{S}(\boldsymbol{\eta})$ is defined as

$$\mathbf{S}(\boldsymbol{\eta}) = \alpha\sqrt{P_T}\mathbf{A}(\mu)\mathbf{C}. \tag{2.22}$$

The FIM $\mathbf{F}(\boldsymbol{\eta})$ is given as

$$\mathbf{F}(\boldsymbol{\eta}) = \begin{bmatrix} \mathrm{F}_{11} & \mathrm{F}_{12} & \mathrm{F}_{13} \\ \mathrm{F}_{12}^{\mathrm{T}} & \mathrm{F}_{22} & \mathrm{F}_{23} \\ \mathrm{F}_{13}^{\mathrm{T}} & \mathrm{F}_{23}^{\mathrm{T}} & \mathrm{F}_{33} \end{bmatrix}. \tag{2.23}$$

The variance var, of the estimation error for each parameter can be lower bounded by the diagonal elements of the inverse of the FIM

$$\mathrm{var}(\hat{\eta}_i) \geq \left[\mathbf{F}^{-1}(\boldsymbol{\eta})\right]_{ii} \tag{2.24}$$

The bound on the error is defined as,

$$\sqrt{CRLB(\hat{\eta}_i)} = \sqrt{[\mathbf{F}^{-1}(\boldsymbol{\eta})]_{ii}}. \tag{2.25}$$

The entries of the $\mathbf{F}(\boldsymbol{\eta})$ are derived in Appendix A.2.

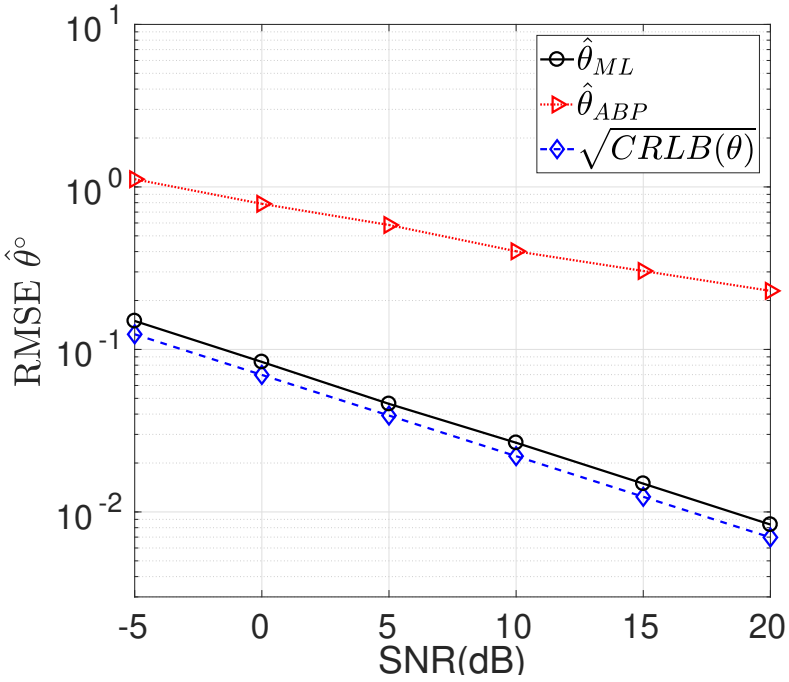


Figure 2.6 – Performance comparison of the proposed ML estimator with ABP method for AoD $\hat{\theta}$ estimation assuming LOS path only.

2.3.4 Numerical Results

In this section, the proposed ML estimator is analyzed and assessed by computer simulations. We adopt the root mean square error (RMSE) using 10000 channel realizations while assessed with the CRLB for statistical analysis as the CRLB sets a benchmark for any unbiased estimator. The average CRLB is calculated with respect to all channel realizations. We also compare our proposed ML estimator with the state-of-the-art approach, i.e., ABP method [26].

The BS is equipped with partially connected Butler matrix (PCBM) approach where each sub-array is controlled by a BM which excites a ULA with $M = 16$ elements, having inter-element spacing $d = \lambda/2$. Moreover, one subarray is used to estimate the channel parameters. The AoD's θ are uniformly generated from a sector of cell as $\theta^\circ \sim \mathrm{U}\left(-60^\circ, +60^\circ\right)$ [47]. The SNR

$$\text{SNR} = 10\log_{10}\frac{P_T|\alpha|^2}{\sigma_n^2}. \tag{2.26}$$

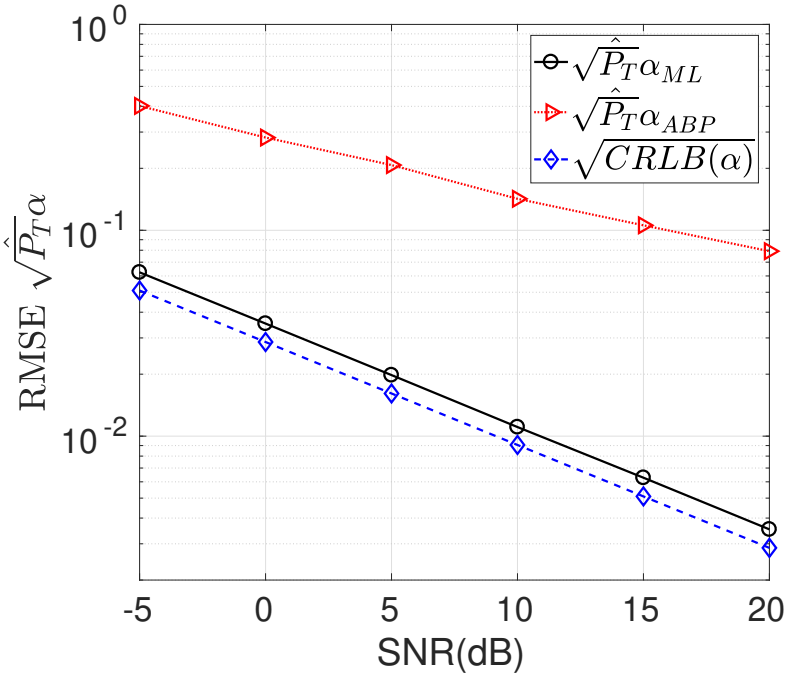


Figure 2.7 – Performance comparison the proposed ML estimator with ABP method for AoD $(\sqrt{\hat{P}_T}\alpha)$ estimation assuming LOS path only.

The RMSE for $\hat{\theta}$ and $(\sqrt{\hat{P}_T}\alpha)$ is

$$\text{RMSE}(\hat{\theta}) = \sqrt{\text{E}[|\theta - \hat{\theta}|^2]} \tag{2.27}$$

$$\text{RMSE}(\sqrt{\hat{P}_T}\alpha) = \sqrt{\text{E}\left[\left|\frac{\sqrt{P_T}\alpha - \sqrt{\hat{P}_T}\alpha}{\sqrt{P_T}\alpha}\right|^2\right]}. \tag{2.28}$$

The error on the bound is calculated as

$$\sqrt{CRLB(\hat{\eta}_i)} = \sqrt{[\mathbf{F}^{-1}(\boldsymbol{\eta})]_{ii}}. \tag{2.29}$$

512 QPSK symbols are used as pilot overhead for parameter estimation. In the ABP method, the channel is probed with multiple beam pairs to estimate the parameters. The delta δ spacing which is important for the design of beam pairs is given as $\delta = \frac{2m\pi}{M}$, where $m = 1, \ldots, \frac{M}{4}$. For choosing $m = 1$ we get $\delta = \frac{\pi}{8}$ which leads to 16 beam pairs $(1,3), (2,4), \ldots, (15,1), (16,2)$. Note that for ABP method, some *a priori* knowledge of AoD has been assumed.

Figure 2.6, shows the performance of the proposed ML estimator and ABP method for frequency flat fading channel for the spatial frequency $\hat{\mu}$, where our proposed ML method throughout outperforms the state-of-the-art ABP approach. To assess the performance of our ML method, we compare with CRLB. The ML estimates approaches nearly the theoretical bound.

Similarly, the Figure 2.7 shows the performance comparison of the proposed ML estimator and the ABP method for the complex path gain $(\sqrt{\hat{P}_T}\alpha)$. Our proposed ML estimator

shows a remarkable performance as compared to the ABP method. Furthermore, the proposed ML estimator approaches the theoretical CRLB.

2.4 Two-stage Estimation Algorithm for Frequency Selective Fading Channel

In this section, a two-stage estimation algorithm is proposed where in the first stage the coarse estimation of the channel parameters is performed and a subsequent search of the channel parameters by approximating the ML estimators using SAGE in the second stage.

2.4.1 System Model

We consider a multiuser downlink scenario assuming a frequency selective channel with R paths, where the BS is equipped with N transmit antennas grouped in N_{RF} sub-arrays having M antennas at each sub-array, i.e., $N = N_{RF} \times M$. The PCBM hybrid beamforming architecture is assumed as shown in Figure 2.3. The received vector per one sub-array by probing one DFT beam with the corresponding specific CAZAC sequence can be shown as

$$\mathbf{y}_k^{\mathrm{T}} = \sqrt{P_T} \sum_{r=1}^{R} \alpha_r \, \mathbf{a}^{\mathrm{H}}(\mu_r)\mathbf{w}(\Phi_k)\mathbf{c}_k^{\mathrm{T}}(\tau_r) + \mathbf{n}_k^{\mathrm{T}} \in \mathbb{C}^{1\times L}, \tag{2.30}$$

P_T is the transmit power, μ_r is the spatial frequency of each multipath $\mu_r = 2\pi\frac{d}{\lambda}\sin\theta_r$, where θ_r is AoD for each path-r, α_r is the complex path gain, τ_r is the delay of each path-r, $\mathbf{a}^*(\mu_r)$ is the channel steering vector assuming ULA for each path-r and $\mathbf{w}(\Phi_k)$ is the DFT beamforming vector having a specific one-to-one correspondence with the CAZAC sequence $\mathbf{c}_k$ composed of QPSK symbols. $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}_{L\times 1}, \sigma_n^2\mathbf{I}_L)$, where σ_n^2 is the noise variance. Now by probing the channel using one sub-array with all the available M DFT beamforming vectors in a round-robin fashion with the corresponding CAZAC sequence $\mathbf{c}_k$ and collect all the received vectors $\mathbf{y}_k$ in the received matrix $\mathbf{Y}$ as

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_0^{\mathrm{T}} \\ \mathbf{y}_1^{\mathrm{T}} \\ \vdots \\ \mathbf{y}_{M-1}^{\mathrm{T}} \end{bmatrix} = \sqrt{P_T} \sum_{r=1}^{R} \alpha_r \mathbf{A}(\mu_r) \underbrace{\begin{bmatrix} \mathbf{c}_0^{\mathrm{T}}(\tau_r) \\ \mathbf{c}_1^{\mathrm{T}}(\tau_r) \\ \vdots \\ \mathbf{c}_{M-1}^{\mathrm{T}}(\tau_r) \end{bmatrix}}_{=\mathbf{C}(\tau_r)} + \underbrace{\begin{bmatrix} \mathbf{n}_0^{\mathrm{T}} \\ \mathbf{n}_1^{\mathrm{T}} \\ \vdots \\ \mathbf{n}_{M-1}^{\mathrm{T}} \end{bmatrix}}_{=\mathbf{N}} \in \mathbb{C}^{M\times L}, \tag{2.31}$$

where $\mathbf{A}(\mu_r) = \mathrm{diag}\{\mathbf{a}^{\mathrm{H}}(\mu_r)\mathbf{w}(\Phi_k)\}_{k=0}^{M-1}$ and the noise covariance matrix $\mathbf{R} = \mathrm{E}\left[\mathrm{vec}\{\mathbf{N}\}\ \mathrm{vec}\{\mathbf{N}\}^{\mathrm{H}}\right] = \sigma_n^2\mathbf{I}_{ML}$.

However, we can also use multiple sub-arrays for channel probing, which will decrease the time for channel probing at the cost of degradation in performance as the total transmit power P_T will be equally divided across multiple sub-arrays.

2.4.2 PREIDG based coarse estimation

In the two-stage estimation algorithm, the first stage is responsible for the coarse estimate of the parameters achieved using PREIDG algorithm. Using M beamforming vectors $\mathbf{w}(\Phi_k), k = 0, \ldots, M-1$ with the corresponding CAZAC sequences $\mathbf{c}_k, k = 0, \ldots, M-1$, where the length of the CAZAC sequence is set to $L = M$. This way the UE observes M consecutive sequences $\mathbf{y}_k$ (2.30) and multiplies these with the already stored CAZAC sequences $\mathbf{c}_k^*$. This can be cast in matrix form as

$$\mathbf{Z} = \mathbf{Y}\mathbf{C}^{\mathrm{H}}(0) = \sqrt{P_T}\sum_{r=1}^{R}\alpha_r\mathbf{A}(\mu_r)\mathbf{C}(\tau_r)\mathbf{C}^{\mathrm{H}}(0) + \mathbf{N}\mathbf{C}^{\mathrm{H}}(0). \tag{2.32}$$

Now for simplicity, let us assume that each of the R AoD's, μ_r (spatial frequency) is equal to one of the Φ_k (beamforming angle) ends up with,

$$\begin{aligned}\mathbf{A}(\mu_r) &= \operatorname{diag}\left\{\mathbf{a}^{\mathrm{H}}(\mu_r)\mathbf{w}(\Phi_k)\right\}_{k=0}^{M-1}\Bigg|_{\mu_r=\Phi_{k_r}} \\ &= \sqrt{M}\,\mathbf{e}_{k_r+1}\mathbf{e}_{k_r+1}^{\mathrm{T}},\end{aligned} \tag{2.33}$$

where $\mathbf{e}_{k_r+1}$ is the M-dimensional canonic unit vector. Furthermore, assume that each path-r has delay which is an integer multiple of the symbol period as

$$\mathbf{C}(\tau_r)\mathbf{C}^{\mathrm{H}}(0)\Big|_{\tau_r=i_r} = M\mathbf{P}_{i_r}, \tag{2.34}$$

where $\mathbf{P}_{i_r} \in \mathbb{R}^{M\times M}$ is the permutation matrix with the following form

$$\begin{aligned}\mathbf{P}_{i_r} &= \sum_{j=1}^{M} \mathbf{e}_j \mathbf{e}_{j+i_r}^{\mathrm{T}} \\ &= \begin{bmatrix} 0 & \dots & 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \dots & \vdots & \vdots & \dots & \ddots & \dots & \vdots \\ \vdots & \dots & \vdots & \vdots & \dots & \dots & \ddots & \vdots \\ 0 & \dots & \vdots & \vdots & \dots & \dots & \dots & 1 \\ 1 & \dots & \vdots & \vdots & \dots & \dots & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \dots & \dots & \dots & \vdots \\ 0 & \dots & 1 & \dots & \dots & \dots & \dots & 0 \end{bmatrix},\end{aligned} \tag{2.35}$$

with $\mathbf{P}_0 = M\mathbf{I}_M$. The matrix $\mathbf{Z}$ consequently is written as

$$\begin{aligned}\mathbf{Z} &= \sqrt{P_T} M \sqrt{M} \sum_{r=1}^{R} \alpha_r \left(\mathbf{e}_{k_r+1} \mathbf{e}_{k_r+1}^{\mathrm{T}} \sum_{j=1}^{M} \mathbf{e}_j \mathbf{e}_{j+i_r}^{\mathrm{T}} \right) + \mathbf{N}\mathbf{C}^{\mathrm{H}}(0) \\ &= \sqrt{P_T} M \sqrt{M} \sum_{r=1}^{R} \alpha_r \mathbf{S}_r + \mathbf{N}\mathbf{C}^{\mathrm{H}}(0).\end{aligned} \tag{2.36}$$

$\mathbf{S}_r$ is a matrix having only the entry in the $k_r + 1$ row and $\mod\ (k_r + i_r + 1, M)$ column equal to one, while rest of all other entries are zero. Therefore, the integer k_r and i_r identify the μ_r and τ_r of the path-r which is on the grid of the DFT beamforming as well of the symbol timing. Hence, the generalized post correlation matrix power $\mathbf{P}$ can be written as

$$\mathbf{P} = \mathbb{E}\left[\mathbf{Z} \odot \mathbf{Z}^*\right] = P_T M^3 \sum_{r=1}^{R} |\alpha_r|^2 \mathbf{S}_r \odot \mathbf{S}_r^* + \sigma_n^2 M \mathbf{1}_M. \tag{2.37}$$

Where the expectation $\mathbb{E}$ is calculated by transmitting multiple copies of the CAZAC sequences. The post-correlation receive SNR is improved by a process gain of M by simple correlation with the CAZAC sequence and further, get an additional M antenna array gain. In a practical scenario, the matrix $\mathbf{S}_r$ will not be strictly sparse having only one non-zero entry. But the power matrix $\mathbf{P}$ at the UE will still have some useful information about μ_r and τ_r. Exploiting the power matrix $\mathbf{P}$ by searching main and wrap-around diagonals $\mathbf{p}_i, i = 1, \dots, M$ of $\mathbf{P}$ for the maximum power as shown in Figure 2.8 for $M = 4$. The searching of warp-around diagonals can be done along $\mathbf{p}_i$, which is given as

$$\mathbf{p}_i^{\mathrm{T}} = \left[p_{1,i}\ ,\ p_{2,\mathrm{mod}(i,M)+1}, \dots, p_{M,\mathrm{mod}(i+M-2,M)+1}\right] \in \mathbb{R}^M, \tag{2.38}$$

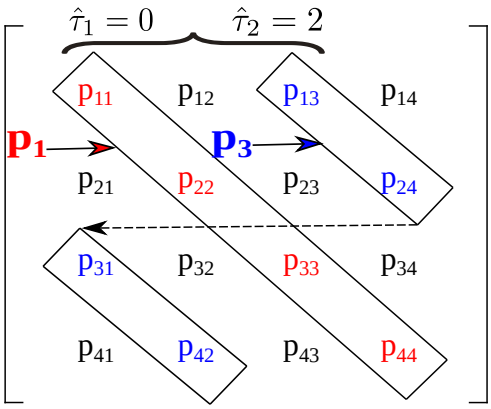


Figure 2.8 – An example of integer delay estimation $\hat{\tau}_i$ using power matrix $\mathbf{P}$ as in (2.37) where $\mathbf{p}_1$ and $\mathbf{p}_3$ are the two diagonals, for which (2.39) is fulfilled.

where $i = 1, \ldots, M$. For $i = 1$ we have $\mathbf{p}_1$, which is the main diagonal of power matrix $\mathbf{P}$. We search all the diagonals $i = 1, \ldots, M$, whether the maximum power exceeds the threshold G as

$$\max_{k=1,\ldots,M} \left(p_{k,\text{mod}(i+(k-2),M)+1}\right) \geq G, \;\; i = 1, \ldots, M. \tag{2.39}$$

G should be heuristically chosen to make sure that the received signal is above the noise floor which in our case is $\sigma_n^2 M$. The coarse estimates for the integer delay τ_{i_r} for path-r is shown in Figure 2.8

$$\hat{\tau}_{i_r} = i_r - 1. \tag{2.40}$$

For the main diagonal $\mathbf{p}_1$, we always assume LOS path, where the delay $\tau_i = i_1 - 1\big|_{i_1=1} = 0$. The number of diagonals fulfilling (2.39) results in the coarse estimation of the model order $\hat{R}$ with the corresponding coarse estimates of $\hat{\tau_{i_r}}$, where we drop the index r for notational convenience.

As we know, the actual spatial frequency for each path-r lies somewhat in between two spatial frequencies $\Phi_k < \mu_r < \Phi_{k+1}$, which ends up with two significant received powers denoted by P_k and P_{k+1} as shown in Figure 2.9.

LUT: To have a coarse estimate of $\hat{\mu}_r$, we construct a LUT which is necessary for the linear interpolation having $D + 1$ spatial frequencies μ_d, generated as

$$\mu_d = \Phi_k + d\Delta_\mu, \;\; d = 0 \ldots, D \tag{2.41}$$

$$\Delta_\mu = \frac{\Phi_{k+1} - \Phi_k}{D} = \frac{2\pi}{MD}. \tag{2.42}$$

Now we compute the hypothetical noise free normalized power based on these μ_d, which is given as

$$P_{k,d} = \left|\mathbf{a}^{\text{H}}(\mu_d)\mathbf{w}(\Phi_k)\right|^2 \tag{2.43}$$

$$P_{k+1,d} = \left|\mathbf{a}^{\text{H}}(\mu_d)\mathbf{w}(\Phi_{k+1})\right|^2. \tag{2.44}$$

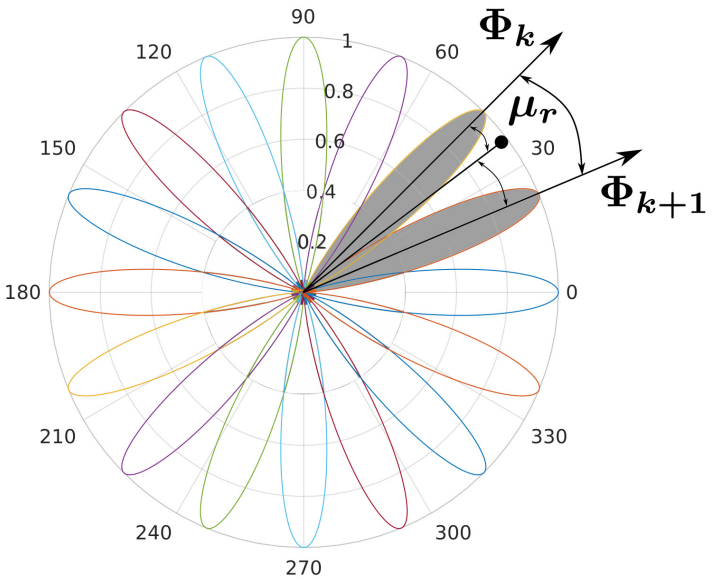


Figure 2.9 – An example of coarse estimation using DFT matrix based beamforming.

where $\mathbf{a}(\mu_d)$ is the channel steering vector and defined as,

$$\mathbf{a}(\mu_d) = \left[1, e^{-j\mu_d}, \ldots, e^{-j(M-1)\mu_d}\right]^{\mathrm{T}} \in \mathbb{C}^{M\times 1}. \tag{2.45}$$

Calculating the ratio of the hypothetical powers as

$$\Delta_d = \sqrt{\frac{P_{k,d}}{P_{k+1,d}}}. \tag{2.46}$$

Since hypothetical powers $P_{k,d}$ and $P_{k+1,d}$ are independent of k, means the difference between any two consecutive beams is the same, therefore we just need $D+1$ ratios to fill into the so-called LUT, respectively. Note that, the LUT once generated is fixed for any coarse estimation of the spatial frequency μ_r.

Coarse estimation of $\hat{\mu}_r$**:** Now to estimate $\hat{\mu}_r$ based on the setup LUT, we choose those indices d and $d+1$ in the LUT, where the two corresponding ratios Δ_d and Δ_{d+1} are closest to the ratio, which the UE has calculated in (2.47)

$$\Delta = \sqrt{\frac{P_k}{P_{k+1}}} \tag{2.47}$$

The coarse estimate of $\hat{\mu}_r$ is calculated using linear interpolation as

$$\hat{\mu}_r = \mu_d + b\Delta_\mu = \Phi_k + \Delta_\mu(d+b). \tag{2.48}$$

where the constant b is given as

$$b = \frac{\Delta_d - \Delta}{\Delta_d - \Delta_{d+1}}. \tag{2.49}$$

Algorithm 2.1: Propose coarse estimation algorithm for estimation of $\hat{\theta}_r$.

1 **Require:** $\mathbf{Y}$ (2.31) ;
2 The UE received $\mathbf{Y}$ and get $\mathbf{p}_i$ for each path r (2.37);
3 Calculate Δ as in (2.47);
4 Find d such that $\Delta_d \geq \Delta \geq \Delta_{d+1}$;
5 Calculate constant b as in (2.49);
6 **Return** $\hat{\mu}_r$ and $\hat{\theta}_r$ as in (2.48) and (2.50).;

Finally, the conversion from spatial frequency $\hat{\mu}_r$ to azimuth AoD $\hat{\theta}_r$

$$\hat{\theta}_r = \begin{cases} \arcsin(\frac{\hat{\mu}_r}{\pi}),\ 0 \leq \hat{\mu}_r \leq \pi, \\ \arcsin(\frac{\hat{\mu}_r - 2\pi}{\pi}),\ \pi < \hat{\mu}_r \leq 2\pi. \end{cases} \tag{2.50}$$

The interpolation between Φ_k and Φ_{k+1} might go wrong in cases where the received signal level is weak and let say μ_r is very close to the beamforming angle Φ_k. In this case, P_k might be quite large, but P_{k+1} might be close to the noise floor. Therefore, for the receiver, it is difficult to decide whether P_{k+1} or P_{k-1} is the second largest power to be used in interpolation due to being masked by noise. Hence, we check whether

$$|P_{k+1} - P_{k-1}| \leq \frac{\sigma_n^2}{v}. \tag{2.51}$$

If (2.51) is satisfied, then it is not worthwhile to interpolate but simply choose $\hat{\mu}_r = \Phi_k$. We heuristically chosen $v = 3$ in our numerical experiments. The complete method is shown in Algorithm 2.1

After estimated all AoDs, the model order estimation may be refined, because of the integer estimation of the delays. One non-integer delay may have lead to two adjacent integer delays, and both of them will have the same AoD estimate. If this occurs, we drop one of the two delays. Furthermore, we can also use a model order detection algorithm given in [48] before feeding the coarse parameters to the SAGE algorithm for the refinement of parameters.

There is still the question, how well this coarse estimation is capable of resolving two paths, which are close to each other in the spatial frequency domain and in the delay domain. If there is a delay difference between the two paths of approximately one symbol period or more, then the spatial frequencies can be arbitrarily close and the paths can still be separated. On the other hand, if the difference of the spatial frequencies is at least $\frac{2\pi}{M}$ (the difference between two adjacent beam patterns), then the two paths could be resolved even if the delay difference is arbitrarily small. Therefore, the proposed coarse estimation scheme is of high resolution in either of the spatial frequency domain or in the delay domain, but not in both simultaneously.

2.4.3 ML Estimation Using SAGE

Let us assume a random matrix $\mathbf{Y}$ which has a multivariate complex Gaussian pdf which is parametrized by the unknown channel parameter vector $\boldsymbol{\eta}$ as

$$\boldsymbol{\eta} = \left[\sqrt{P_T}\mathrm{Re}\{\boldsymbol{\alpha}\}^{\mathrm{T}}, \sqrt{P_T}\mathrm{Im}\{\boldsymbol{\alpha}\}^{\mathrm{T}}, \boldsymbol{\mu}^{\mathrm{T}}, \boldsymbol{\tau}^{\mathrm{T}}\right]^{\mathrm{T}}. \tag{2.52}$$

where

$$\sqrt{P_T}\mathrm{Re}\{\boldsymbol{\alpha}\} = \left[\sqrt{P_T}\mathrm{Re}\{\alpha_1\}, \ldots, \sqrt{P_T}\mathrm{Re}\{\alpha_r\}, \ldots, \sqrt{P_T}\mathrm{Re}\{\alpha_R\}\right]^{\mathrm{T}}, \tag{2.53}$$

$$\sqrt{P_T}\mathrm{Im}\{\boldsymbol{\alpha}\} = \left[\sqrt{P_T}\mathrm{Im}\{\alpha_1\}, \ldots, \sqrt{P_T}\mathrm{Im}\{\alpha_r\}, \ldots, \sqrt{P_T}\mathrm{Im}\{\alpha_R\}\right]^{\mathrm{T}}, \tag{2.54}$$

$$\boldsymbol{\mu} = \left[\mu_1, \ldots, \mu_r, \ldots, \mu_R\right]^{\mathrm{T}}, \tag{2.55}$$

$$\boldsymbol{\tau} = \left[\tau_1, \ldots, \tau_r, \ldots, \tau_R\right]^{\mathrm{T}}. \tag{2.56}$$

The likelihood function is

$$\mathrm{L}(\mathbf{Y};\boldsymbol{\eta}) = \frac{1}{\pi^{ML}\det\mathbf{R}} \exp\left(-\mathrm{vec}\left\{\mathbf{Y} - \sqrt{P_T}\sum_{r=1}^{R}\alpha_r\mathbf{A}(\mu_r)\mathbf{C}(\tau_r)\right\}^{\mathrm{H}} \mathbf{R}^{-1}\,\mathrm{vec}\left\{\mathbf{Y} - \sqrt{P_T}\sum_{r=1}^{R}\alpha_r\mathbf{A}(\mu_r)\mathbf{C}(\tau_r)\right\}\right). \tag{2.57}$$

The ML estimator is given as

$$\hat{\boldsymbol{\eta}} = \arg\max_{\boldsymbol{\eta}} \mathrm{L}\left(\mathbf{Y};\boldsymbol{\eta}\right), \tag{2.58}$$

whereas no closed form solution can found for the ML estimator because (2.58) is a multidimensional, non-linear optimization problem. We use the SAGE algorithm [49] to solve our non-linear problem. We configure the SAGE algorithm for our scenario and ends up with the following expectation-maximization steps,

Expectation step:

$$\hat{\mathbf{X}}_r = \mathbf{Y} - \sum_{\substack{r'=1\\ r'\neq r}}^{R}\mathbf{S}_{r'}(\hat{\boldsymbol{\eta}}_{r'}). \tag{2.59}$$

where

$$\mathbf{S}_r(\boldsymbol{\eta}_r) = \sqrt{P_T}\alpha_r \mathbf{A}(\mu_r)\mathbf{C}(\tau_r). \tag{2.60}$$

Maximization step:

$$\hat{\tau}_r = \arg\max_{\tau_r} \left\{ \frac{\left| \mathrm{tr}\left\{ \mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\hat{\mu}_r)\hat{\mathbf{X}}_r \right\} \right|^2}{\beta_r \sigma_n^2 \ \mathrm{tr}\{\mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\hat{\mu}_r)\mathbf{A}(\hat{\mu}_r)\mathbf{C}(\tau_r)\}} \right\}, \tag{2.61}$$

$$\hat{\mu}_r = \arg\max_{\mu_r} \left\{ \frac{\left| \mathrm{tr}\left\{ \mathbf{C}^{\mathrm{H}}(\hat{\tau}_r)\mathbf{A}^{\mathrm{H}}(\mu_r)\hat{\mathbf{X}}_r \right\} \right|^2}{\beta_r \sigma_n^2 \ \mathrm{tr}\{\mathbf{C}^{\mathrm{H}}(\hat{\tau}_r)\mathbf{A}^{\mathrm{H}}(\mu_r)\mathbf{A}(\mu_r)\mathbf{C}(\hat{\tau}_r)\}} \right\}. \tag{2.62}$$

Finally, $\hat{\sqrt{P_T}\alpha}$ can be found as

$$\hat{\sqrt{P_T}\alpha}_r = \frac{\mathrm{tr}\left\{ \mathbf{C}^{\mathrm{H}}(\hat{\tau}_r)\mathbf{A}^{\mathrm{H}}(\hat{\mu}_r)\hat{\mathbf{X}}_r \right\}}{\mathrm{tr}\{\mathbf{C}^{\mathrm{H}}(\hat{\tau}_r)\mathbf{A}^{\mathrm{H}}(\hat{\mu}_r)\mathbf{A}(\hat{\mu}_r)\mathbf{C}(\hat{\tau}_r)\}}. \tag{2.63}$$

The derivation is given in Appendix A.3.

2.4.3.1 Convergence analysis of SAGE

The second stage of the two-stage algorithm is to refine the coarse estimate using the ML estimator. To make sure that the SAGE algorithm converges to the global optimum using very few iterations is of practical interest. To achieve this, it is mandatory to start the SAGE algorithm with a good coarse estimate of the parameters achieved using PREIDG algorithm. We use $\hat{\tau}_i$ (2.40), $\hat{\mu}_r$ (2.48) and $\hat{\alpha}_r = 0$ to initialize the SAGE in the second stage of the algorithm. One iteration of the SAGE algorithm is as a full update of the parameter vector $\boldsymbol{\eta}$. The stopping thresholds for the SAGE convergence are given as

$$\mathrm{T}_1 = \frac{|\hat{\mu}_{r_p} - \hat{\mu}_r|}{|\hat{\mu}_r|}, \tag{2.64}$$

$$\mathrm{T}_2 = \frac{|\hat{\tau}_{r_p} - \hat{\tau}_r|}{|\hat{\tau}_r|}, \tag{2.65}$$

$$\mathrm{T}_3 = \frac{|\hat{\sqrt{P_T}\alpha}_{r_p} - \hat{\sqrt{P_T}\alpha}_r|}{|\hat{\sqrt{P_T}\alpha}_r|}, \tag{2.66}$$

where $\hat{\mu}_{r_p}, \hat{\tau}_{r_p}, \hat{\sqrt{P_T}\alpha}_{r_p}$, are the previous estimates of spatial frequency, delay time and complex path gain. The SAGE algorithm approach approximately the global optimum when (2.67) is

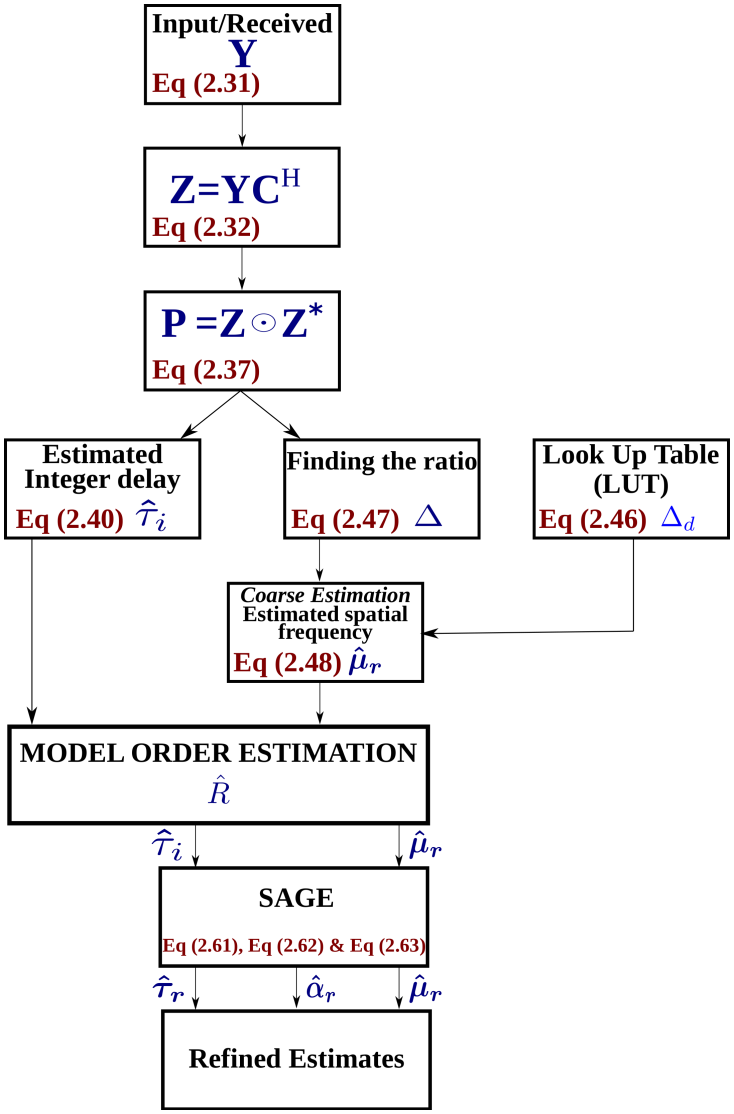


Figure 2.10 – A flowchart for the two-stage estimation algorithm.

fulfilled as

$$\max\{\mathrm{T}_1, \mathrm{T}_2, \mathrm{T}_3\} \leq \Gamma, \tag{2.67}$$

Γ represents the stopping threshold.

The two-stage algorithm is shown in Figure 2.10.

2.4.4 CRLB

In this section, we derived the FIM $\mathbf{F}(\boldsymbol{\eta})$ for providing a theoretical lower bound on the estimation errors of the channel parameters $\boldsymbol{\eta}$ to assess the performance our approximated ML estimates.

Assuming $\hat{\boldsymbol{\eta}}$ as an unbiased estimator of $\boldsymbol{\eta}$, then the variance var of the estimation error can be

lower-bounded by the diagonal elements of the inverse of $\mathbf{F}(\boldsymbol{\eta})$ [46]

$$\mathrm{var}(\hat{\eta}_i) \geq \left[\mathbf{F}^{-1}(\boldsymbol{\eta})\right]_{ii}. \tag{2.68}$$

The FIM for the complex data can be given as

$$[\mathbf{F}(\boldsymbol{\eta})]_{ij} = \frac{2}{\sigma_n^2}\ \mathrm{Re}\ \left\{\mathrm{tr}\left\{\frac{\partial \mathbf{S}^{\mathrm{H}}(\boldsymbol{\eta})}{\partial \eta_i}\ \frac{\partial \mathbf{S}(\boldsymbol{\eta})}{\partial \eta_j}\right\}\right\}, \tag{2.69}$$

where the matrix $\mathbf{S}(\boldsymbol{\eta})$ is

$$\mathbf{S}(\boldsymbol{\eta}) = \sum_{r=1}^{R} \mathbf{S}(\boldsymbol{\eta}_r) = \sqrt{P_T}\ \sum_{r=1}^{R} \alpha_r\, \mathbf{A}(\mu_r)\, \mathbf{C}(\tau_r) \tag{2.70}$$

where

$$\boldsymbol{\eta}_r = \left[\sqrt{P_T}\mathrm{Re}\{\alpha_r\}, \sqrt{P_T}\mathrm{Im}\{\alpha_r\}, \mu_r, \tau_r\right]. \tag{2.71}$$

The FIM $\mathbf{F}(\boldsymbol{\eta})$ can be structured as

$$\mathbf{F}(\boldsymbol{\eta}) = \begin{bmatrix} \mathbf{F}_{\mathrm{Re}\{\boldsymbol{\alpha}\}\mathrm{Re}\{\boldsymbol{\alpha}\}} & \mathbf{F}_{\mathrm{Re}\{\boldsymbol{\alpha}\}\mathrm{Im}\{\boldsymbol{\alpha}\}} & \mathbf{F}_{\mathrm{Re}\{\boldsymbol{\alpha}\}\boldsymbol{\mu}} & \mathbf{F}_{\mathrm{Re}\{\boldsymbol{\alpha}\}\boldsymbol{\tau}} \\ \mathbf{F}^{\mathrm{T}}_{\mathrm{Re}\{\boldsymbol{\alpha}\}\mathrm{Im}\{\boldsymbol{\alpha}\}} & \mathbf{F}_{\mathrm{Im}\{\boldsymbol{\alpha}\}\mathrm{Im}\{\boldsymbol{\alpha}\}} & \mathbf{F}_{\mathrm{Im}\{\boldsymbol{\alpha}\}\boldsymbol{\mu}} & \mathbf{F}_{\mathrm{Im}\{\boldsymbol{\alpha}\}\boldsymbol{\tau}} \\ \mathbf{F}^{\mathrm{T}}_{\mathrm{Re}\{\boldsymbol{\alpha}\}\boldsymbol{\mu}} & \mathbf{F}^{\mathrm{T}}_{\mathrm{Im}\{\boldsymbol{\alpha}\}\boldsymbol{\mu}} & \mathbf{F}_{\boldsymbol{\mu}\boldsymbol{\mu}} & \mathbf{F}_{\boldsymbol{\mu}\boldsymbol{\tau}} \\ \mathbf{F}^{\mathrm{T}}_{\mathrm{Re}\{\boldsymbol{\alpha}\}\boldsymbol{\tau}} & \mathbf{F}^{\mathrm{T}}_{\mathrm{Im}\{\boldsymbol{\alpha}\}\boldsymbol{\tau}} & \mathbf{F}^{\mathrm{T}}_{\boldsymbol{\mu}\boldsymbol{\tau}} & \mathbf{F}_{\boldsymbol{\tau}\boldsymbol{\tau}} \end{bmatrix}. \tag{2.72}$$

The entries of the $\mathbf{F}(\boldsymbol{\eta})$ (2.72) are derived in Appendix A.4.

2.4.5 Numerical Results

In this section, the performance of the proposed two-stage algorithm is analyzed and assessed with CRLB and compared with the ABP method [26] by computer simulations. The transmitter at the BS station uses the PCBM hybrid architecture with the BM used for exciting the ULA with $M = 16$ antennas, using the inter-element spacing $d = \lambda/2$. Furthermore, we assume only one subarray for channel probing in a round-robin fashion. We assume the carrier frequency of the system $f_c = 28\,\mathrm{GHz}$ and bandwidth of the system as $B = 200\,\mathrm{MHz}$ giving a symbol duration $T_s = 5\,\mathrm{ns}$. We use 32 pilot symbols per beam for channel probing which leads to the total number of 512 QPSK symbols as pilot overhead. We model our channel parameters based on [47]. The distance for the LOS among the BS and UE is uniformly distributed as

$$X_{\mathrm{los}} \sim \mathrm{U}(30\,\mathrm{m}, 60\,\mathrm{m}). \tag{2.73}$$

Similarly, the NLOS distance are uniformly distributed as

Table 2.1 – Beamforming angles using one 16×16 BM

S.No	**beamforming vector**	**spatial frequency** Φ_k	**AoD** θ_k°
1	$\mathbf{w}(\Phi_0)$	0	0
2	$\mathbf{w}(\Phi_1)$	$\pi/8$	7.1808
3	$\mathbf{w}(\Phi_2)$	$\pi/4$	14.4775
4	$\mathbf{w}(\Phi_3)$	$3\pi/8$	22.0243
5	$\mathbf{w}(\Phi_4)$	$\pi/2$	30
6	$\mathbf{w}(\Phi_5)$	$5\pi/8$	38.6822
7	$\mathbf{w}(\Phi_6)$	$3\pi/4$	48.5904
8	$\mathbf{w}(\Phi_7)$	$7\pi/8$	61.0450
9	$\mathbf{w}(\Phi_8)$	π	90
10	$\mathbf{w}(\Phi_9)$	$9\pi/8$	-61.0450
11	$\mathbf{w}(\Phi_{10})$	$5\pi/4$	-48.5904
12	$\mathbf{w}(\Phi_{11})$	$11\pi/8$	-38.6822
13	$\mathbf{w}(\Phi_{12})$	$3\pi/2$	-30
14	$\mathbf{w}(\Phi_{13})$	$13\pi/8$	-22.0243
15	$\mathbf{w}(\Phi_{14})$	$7\pi/4$	-14.4775
16	$\mathbf{w}(\Phi_{15})$	$15\pi/8$	-7.1808

$$X_{\text{nlos}} = X_{\text{los}} + \Delta_{\text{nlos}}, \tag{2.74}$$

where Δ_{nlos} is the difference between LOS and NLOS component and is distributed as $\Delta_{\text{nlos}} \sim \mathrm{U}(4.5\,\mathrm{m}, 24\,\mathrm{m})$ which ends up with a delay difference between 3 and 16 symbols. The length of CAZAC sequence $L = 16$ restricts the maximum delay difference to be estimated. The path loss can be given as

$$PL(\mathrm{dB}) = 10\bar{n}\log_{10}\left(\frac{X}{X_0}\right), \tag{2.75}$$

where $\bar{n}$ represents the path loss exponent which is chosen as 2.1 and 2.4 for LOS and NLOS respectively. X_0 is assumed as one meter. γ_r is magnitude of the complex path gain of each path-r and is defined as the ratio of the path loss for LOS (PL_{los}) and NLOS (PL_{nlos}) respectively.

$$\gamma_r = \sqrt{\frac{PL_{\text{los}}}{PL_{\text{nlos}}}}. \tag{2.76}$$

The complex path gain α_r for each path-r can be given as

$$\frac{\alpha_r}{\alpha_1} = \gamma_r e^{j\varphi_r}, \tag{2.77}$$

where α_1 is the complex path gain of the LOS and assumed as one. φ_r is the phase of the complex path coefficient. For every path-r, it is generated as $\varphi_r \sim \mathrm{U}(0, 2\pi)$. Note that the phase of the complex gain φ_r is different for every sub-array at BS while the magnitude γ_r is the same for every sub-array. In this approach, as the BS knows the geometry of the antennas array, it can estimate the phase of the other sub-arrays based on the estimation of the phase φ_1 of the first array.

The AoD's θ_r for LOS and NLOS can be generated simultaneously as being uniformly distributed over one sector of a cell $\theta_r^\circ \sim \mathrm{U}(-60^\circ, +60^\circ)$. The noise variance is assumed as $\sigma_n^2 = 1$. The total number of paths is assumed as $R = 3$. The spatial frequencies and their respective AoD azimuth angles in degrees are given in Table 2.1. The SNR is defined as

$$\text{SNR} = \frac{P_T|\alpha_1|^2}{\sigma_n^2}. \tag{2.78}$$

The RMSE for the channel parameter can be given as

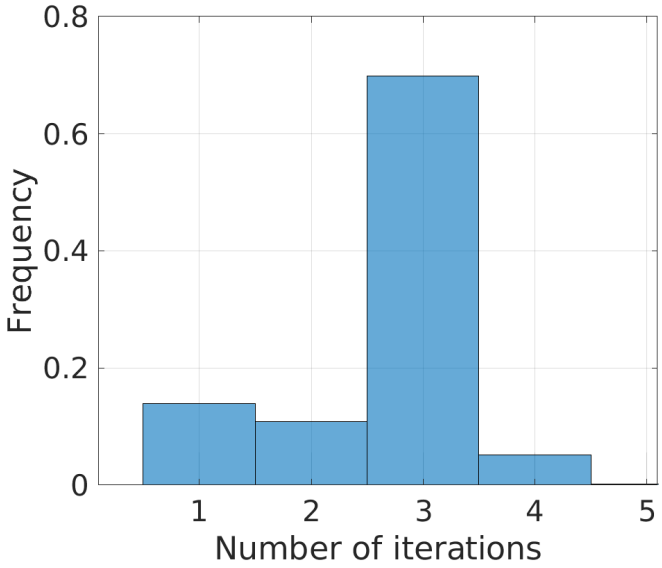


Figure 2.11 – Histogram for the number of iterations of SAGE in the two-stage algorithm.

$$\text{RMSE}(\hat{\theta}_r) = \sqrt{\mathbb{E}\left[|\theta_r - \hat{\theta}_r|^2\right]} \tag{2.79}$$

$$\text{RMSE}(\sqrt{\hat{P_T}}\alpha_r) = \sqrt{\mathbb{E}\left[\left|\frac{\sqrt{P_T}\alpha_r - \sqrt{\hat{P_T}}\alpha_r}{\sqrt{P_T}\alpha_r}\right|^2\right]} \tag{2.80}$$

$$\text{RMSE}(\hat{\tau}_r) = \sqrt{\mathbb{E}\left[|\tau_r - \hat{\tau}_r|^2\right]}. \tag{2.81}$$

The bound on the error is calculated as

$$\sqrt{CRLB(\hat{\eta}_i)} = \sqrt{[\mathbf{F}^{-1}(\boldsymbol{\eta})]_{i\,i}}. \tag{2.82}$$

We use 10000 channel realizations for calculating the RMSE for each parameter at different SNRs and similarly, the CRLB for each parameter is obtained by numerical averaging over 10000 channel scenarios. The length of the look-up/interpolation table for the PREIDG algorithm is assumed as $D = 101$.

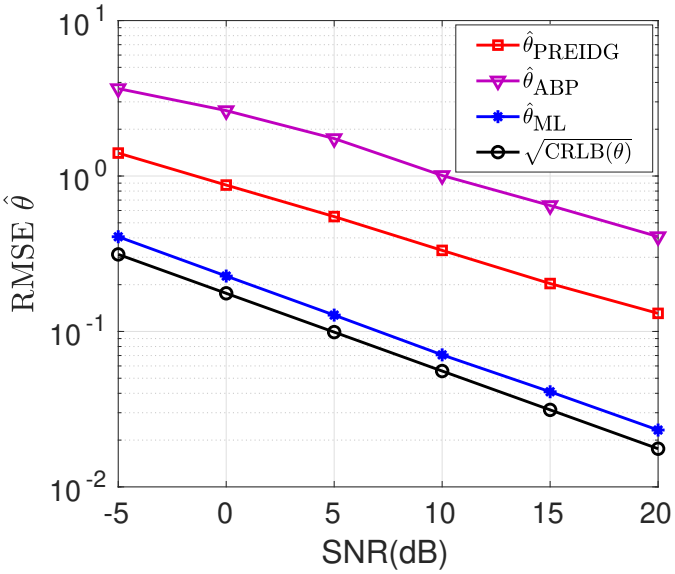


Figure 2.12 – Performance comparison of two-stage and ABP algorithm for LOS AoD, assuming $R = 3$.

To form beam pairs in the ABP method using a DFT matrix, we fixed $\delta = \frac{2m\pi}{M} = \frac{\pi}{8}$, assuming $m = 1$ [26]. The 16 beam pairs can be represented as $(1, 3), (2, 4), \ldots, (15, 1), (16, 2)$. The criteria for choosing the auxiliary beam pair out of all the 16 beam pairs is the one that gives the maximum average power.

The motivation behind the two-stage algorithm is to reduce the number of iterations for the convergence of the SAGE and assure it to achieve the global optimum. For this reason, an additional coarse estimation based on PREIDG is performed before the SAGE algorithm is initialized. Now with initialization of the SAGE algorithm with $\hat{\mu}_r$ (2.48), integer delay $\hat{\tau}_i$ (2.40), $\sqrt{P_T}\alpha_r = 0$, and fixing the threshold $\Gamma = 10^{-3}$, which facilitate the convergence of the SAGE algorithm with maximum number of 4 iterations. In 70% of the channel realizations, SAGE took 3 iterations to converge, which is shown in Figure 2.11. In the low SNR regime, during the coarse estimation, we are also doing the model order estimation due to which the NLOS paths are not always detected and, unable to estimate in every channel realization because of the high path-loss. While in the high SNR regime, we can detect almost all paths.

Figure 2.12 shows the performance of the proposed two-stage algorithm as compared with the ABP method. Simulation results based on 10 thousands of channel realizations show that the proposed PREIDG method performs better than ABP. Furthermore, it shows that the resolution of the PREIDG algorithm improves by increasing SNR. After using coarse estimation based on PREIDG, as an ad-hoc estimation to initialize the SAGE algorithm gives the improved ML performance which nearly satisfies the theoretical bound.

Similarly, the AoDs performance of the two-stage method for NLOS paths are compared with

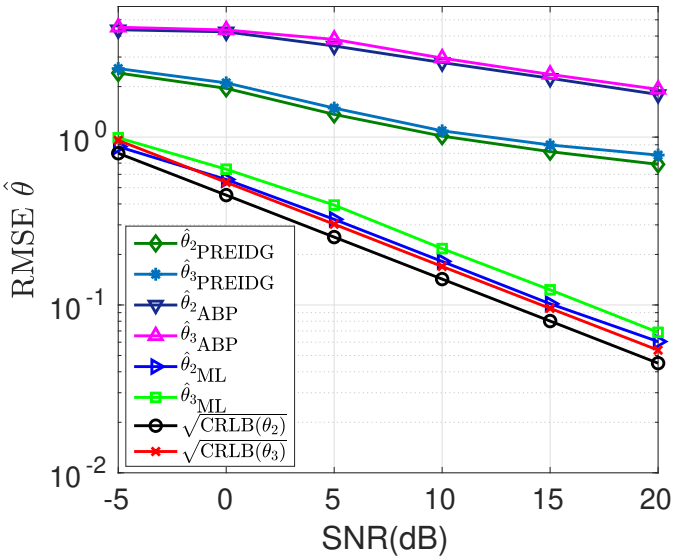


Figure 2.13 – Performance comparison of two-stage and ABP algorithm for NLOS AoD, assuming $R = 3$.

ABP and is assessed with theoretical CRLB in Figure 2.13. The performance of PREIDG still performs better than ABP for NLOS paths. The performance of the ML approach performs better than both PREIDG and ABP method. The ML method approaches the CRLB closely.

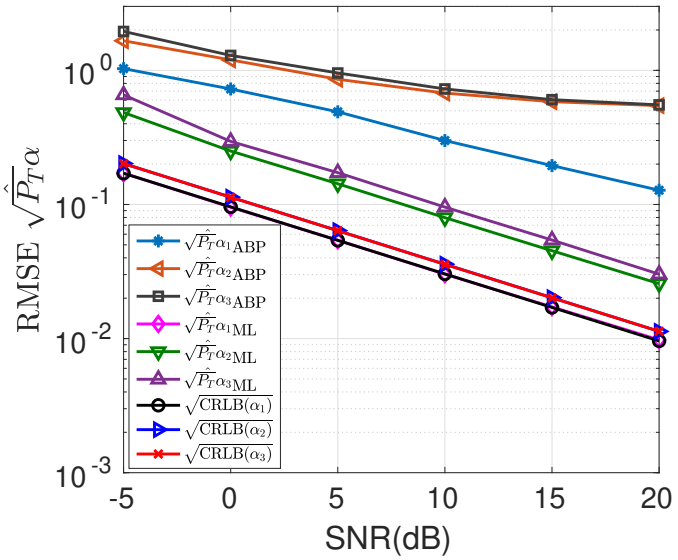


Figure 2.14 – Performance comparison of two-stage and ABP algorithm for α_r.

Figure 2.14 shows the comparison of the two-stage algorithm with the ABP method. The proposed two-stage algorithm shows a better performance for the estimation of the complex path gain $(\sqrt{\hat{P_T}}\alpha_r)$ as compared to the ABP method. In addition, the ML estimate as a result

of the two-stage algorithm for $(\sqrt{\hat{P_T}}\alpha_r)$ approximately approaches the theoretical CRLB. We assume $\alpha_1 = 1$ for LOS path.

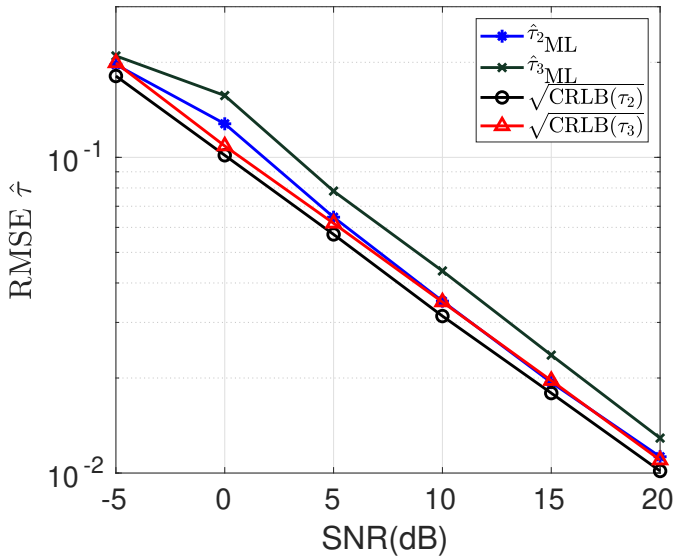


Figure 2.15 – Performance comparison of two-stage algorithm for τ_r

The two-stage estimation algorithm is evaluated for delay time τ_r in fractions of symbol period as shown in Figure 2.15. The result is the output of the two-stage algorithm which approximates ML estimate for τ_r which nearly approaches the theoretical CRLB. The delay time for LOS is assumed as $\tau_1 = 0$. The combining of coarse estimation based on a PREIDG algorithm to initialize the SAGE algorithm enhances the estimation accuracy. The ABP algorithm is unable to estimate the delay time τ_r.

Keep in mind that we can also use ABP method for coarse estimation but it is not guaranteed that the SAGE algorithm will converge to the global optimum. This is because that the ABP method is unable to estimate the time-delay for each path.

3 TWO-DIMENSIONAL CHANNEL PARAMETER ESTIMATION

This chapter introduces the two-dimensional channel parameter estimation (azimuth and elevation angle of departure, complex-path gain and time delay) using a two-stage algorithm. In the first stage, a coarse estimation is based on modified PREIDG using Kronecker products of DFT beams while high-resolution parameter estimation is done using the SAGE algorithm. To assess the performance of the SAGE algorithm, CRLB is derived in addition to the comparison of a modified PREIDG algorithm with the ABP method.

3.1 Overview

To realize two-dimensional beamforming for the next-generation mobile communication systems, an URA is introduced which simplifies multiuser spatial separation in a cell while maximizing the available data rates and minimizing the inter-user interference. Employing large antenna arrays at the BS is a challenge that is tackled by introducing hybrid beamforming with a PCAPS network using Butler matrices (BMs) termed as partially connected Butler matrices (PCBMS) approach in the analog domain to ensure energy efficient implementation [50, 51, 45, 52]. The idea is to extend the ULA excited by a single BM as discussed in [50, 51] to a two-dimensional URA, which is of more practical interest. We propose a new way of combing BMs to excite an URA as shown in Figure 3.1. In this two-dimensional network of BMs, the resultant beamforming vector is the Kronecker product of the columns of DFT

$$\mathbf{w}(\delta_p, \nu_q) = \mathbf{w}_h(\delta_p) \otimes \mathbf{w}_v(\nu_q) \in \mathbb{C}^{M\times 1}, \tag{3.1}$$

where $\mathbf{w}\left(\delta_p, \nu_q\right)$ is the resultant beamforming vector for probing the channel. δ_p is the DFT spatial frequency of the DFT matrix controlling the horizontal antenna elements, and $\mathbf{w}_h(\delta_p)$ is defined as

$$\mathbf{w}_h(\delta_p) = \frac{1}{\sqrt{M_h}}\left[1, e^{-j\delta_p}, \ldots, e^{-j(M_h-1)\delta_p}\right]^{\mathrm{T}} \in \mathbb{C}^{M_h\times 1}, \tag{3.2}$$

where $\delta_p = \frac{2\pi}{M_h}p$, $p = 0, \ldots, M_h - 1$. M_h is the number of antenna elements on the horizontal axis. ν_q is the DFT spatial frequency of the DFT matrix exciting the vertical antenna elements in the URA, and $\mathbf{w}_v(\nu_q)$ is defined as

$$\mathbf{w}_v(\nu_q) = \frac{1}{\sqrt{M_v}}\left[1, e^{-j\nu_q}, \ldots, e^{-j(M_v-1)\nu_q}\right]^{\mathrm{T}} \in \mathbb{C}^{M_v\times 1}, \tag{3.3}$$

$\nu_q = \frac{2\pi}{M_v}q$, $q = 0, \ldots, M_v - 1$. M_v represents the number of antenna elements on vertical axis.

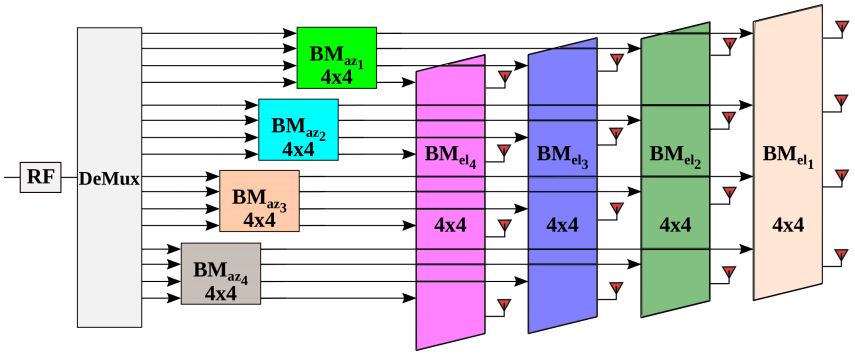


Figure 3.1 – An architecture of BMs exciting two-dimensional uniform rectangular array.

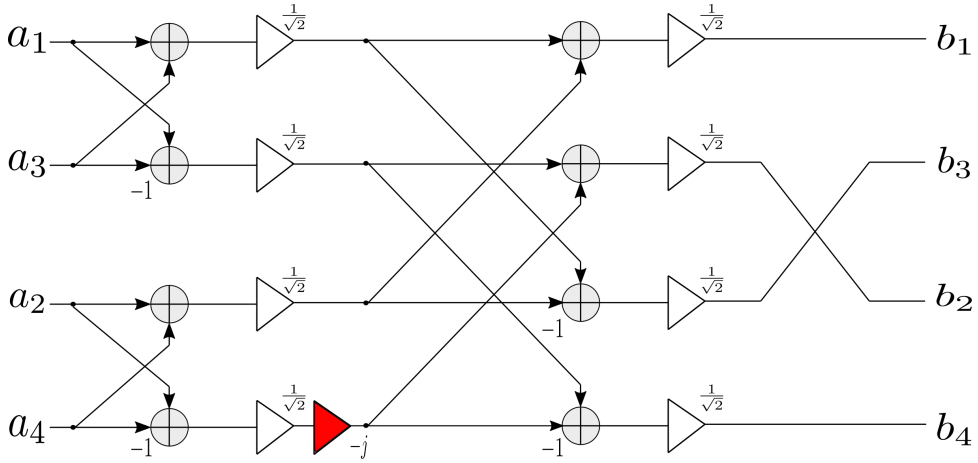


Figure 3.2 – A 4-point DFT building block.

3.1.1 Construction of 4×4 *BM*

The fast Fourier transform (FFT)-signal-flow graph (SFG) for a 4×4 DFT matrix multiplication is accomplished with eight complex additions and one complex multiplication as shown in Figure 3.2, with the following expressions,

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -j & -1 & j \\ 1 & -1 & 1 & -1 \\ 1 & j & -1 & -j \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix}. \tag{3.4}$$

There is a total of four butterflies in the construction of 4-point DFT matrix. To design 4×4 microwave circuit, i.e., BM using 90° hybrid coupler as shown in Figure 2.4, we need to replace the one butterfly with one 90° hybrid coupler as given in (2.2) with two PSs and one additional -90° fixed PS as shown in Figure 3.3.

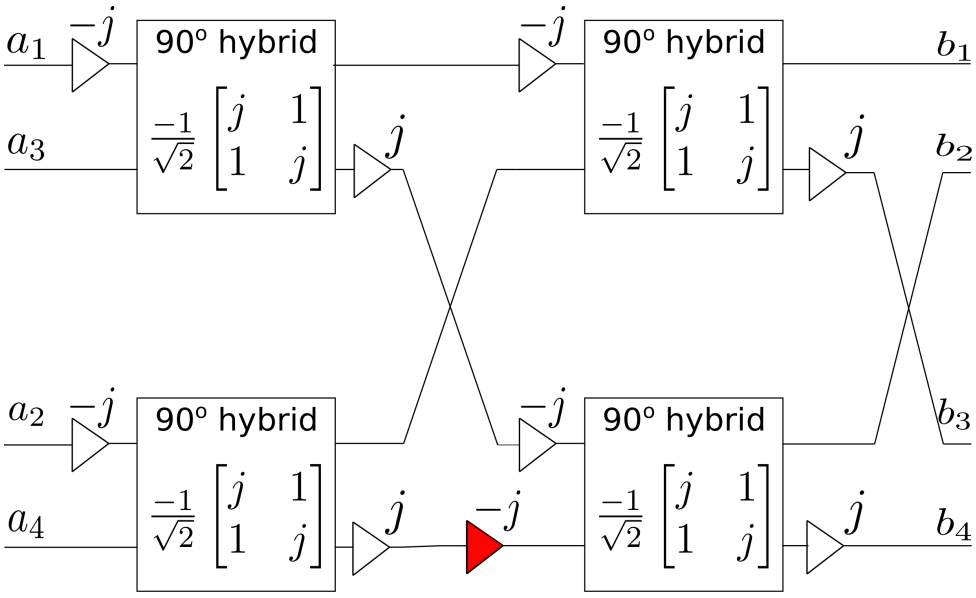


Figure 3.3 – A 4×4 BM using 90° hybrid couplers.

3.2 Contributions

The contribution and the organization of the chapter is as

- **System Model:** In this section, the system model is based on a two-dimensional URA excited by the network BMs in the analog domain.
- **Coarse estimation based on modified PREIDG:**
  In this section, the coarse estimation is performed based on the modified PREIDG.
- **SAGE:**
  We derive and configure SAGE algorithm for the non-linear problem to achieve high-resolution parameter estimation by initializing with coarse estimates achieved using modified PREIDG.
- **CRLB:**
  We derive the CRLB for high-resolution channel parameter estimation considering MISO system where the BS is equipped with an URA and the UEs have a single antenna each.
- **Simulation results:** In this section, we show extensive simulations to evaluate the performance of the proposed modified PREIDG and compare it with the ABP method [27]. We further show that, initializing the SAGE algorithm with the coarse estimates achieved using modified PREIDG algorithm can nearly approache the CRLB.

  The author's research contributions include:

  1. Design of the two-dimensional beamformer with hardware constraints. We introduce new way of combining BMs to excite the URA. The new architecture is energy-

efficient and has less implementation complexity in the analog domain.

2. A novel two-stage algorithm for the two-dimensional parameter estimation is proposed, where an efficient way of coarse estimation is achieved based on modified PREIDG algorithm followed by the refinement stage done by the SAGE algorithm.
3. Derivation of the CRLB for the assessment of the SAGE algorithm. The modified PREIDG is compared with the ABP method [27].

3.3 Two-Dimensional Parameter Estimation Algorithm for Frequency Selective Channel

In the section, a novel two-dimensional two-stage algorithm is proposed and discussed. A coarse estimation based on the modified PREIDG is used to the SAGE algorithm to achieve high-resolution parameter estimation.

3.3.1 System Model

We consider a single-cell downlink scenario for a frequency selective channel considering a MISO system where the BS is equipped with N antennas. The BS employs a PCBMS approach [50, 51] where each sub-array is composed of M antennas satisfying $N = N_{RF} \times M$, controlled by N_{RF} chains. The two-dimensional mth beamforming vector $\mathbf{w}_m\left(\delta_p, \nu_q\right)$ is used to probe the channel with a specific CAZAC [53, 54, 55] sequence, leading to the resulting received vector $\mathbf{y}_m \in \mathbb{C}^{L\times 1}$ per one sub-array

$$\mathbf{y}_m^{\mathrm{T}} = \sqrt{P_T}\sum_{r=1}^{R}\alpha_r \mathbf{a}^{\mathrm{H}}(\phi_r,\theta_r)\mathbf{w}_m(\delta_p,\nu_q)\mathbf{c}_m^{\mathrm{T}}(\tau_r) + \mathbf{n}_m^{\mathrm{T}}, \tag{3.5}$$

where $m = q + pM_v$, $p \in \{0,\ldots,M_h-1\}$, $q \in \{0,\ldots,M_v-1\}$, and $m \in \{0,\ldots,M-1\}$ given that $M = M_hM_v$. P_T is the transmit power, α_r is the complex gain of each path r, $\mathbf{c}_m$ is the specific CAZAC sequence with length L having one-to-one correspondence with the two-dimensional beamforming vector $\mathbf{w}_m\left(\delta_p, \nu_q\right)$, $\mathbf{n}_m \sim \mathcal{CN}(\mathbf{0}_{L\times 1}, \sigma_n^2\mathbf{I}_L)$ is the noise vector, where σ_n^2 is the noise variance, ϕ_r is the azimuth AoD, θ_r is the elevation AoD, $\mathbf{a}(\phi_r,\theta_r)$ is the two-dimensional steering vector of a URA in the y-z-plane with inter-elements distance as $d_h = d_v = \lambda/2$ as,

$$a_m(\phi_r,\theta_r) = e^{-j\pi[p\sin\theta_r\sin\phi_r + q\cos\theta_r]} \tag{3.6}$$

The URA has M_h antenna elements in the y direction and M_v elements in the z direction with $M = M_hM_v$ as shown in Figure 3.4. Now, let us consider the two-dimensional steering

vector $\mathbf{a}\left(\phi_r, \theta_r\right)$ given in (3.6) and define the spatial frequencies as $\mu_r = \pi \sin\theta_r \sin\phi_r$ and $\psi_r = \pi \cos\theta_r$ respectively. Therefore, the steering vector can be re-written as

$$\mathbf{a}(\mu_r, \psi_r) = \left[a_0(\mu_r, \psi_r), a_1(\mu_r, \psi_r), \ldots, a_{M-1}(\mu_r, \psi_r)\right]^{\mathrm{T}}. \tag{3.7}$$

Furthermore, (3.7) can also be re-written as the Kronecker product using the following steering vectors

$$\boldsymbol{a}_h(\mu_r) = \left[1, e^{-j\mu_r}, \ldots, e^{-j(Mv-1)\mu_r}\right] \in \mathbb{C}^{Mv\times 1}, \tag{3.8}$$

and

$$\boldsymbol{a}_v(\psi_r) = \left[1, e^{-j\psi_r}, \ldots, e^{-j(Mh-1)\psi_r}\right] \in \mathbb{C}^{Mh\times 1}. \tag{3.9}$$

Hence, the channel steering vector $\mathbf{a}\left(\mu_r, \psi_r\right)$ can also be written as

$$\boldsymbol{a}(\mu_r, \psi_r) = \boldsymbol{a}_h(\mu_r) \otimes \boldsymbol{a}_v(\psi_r) \in \mathbb{C}^{M\times 1}. \tag{3.10}$$

The CAZAC sequence specific for each beamforming vector is $\mathbf{c}_m$, where each symbol c_0 is constructed as

$$c(n) = e^{\left(j\frac{2\pi}{\sqrt{L}}\left(\mathrm{mod}\left\{n,\sqrt{L}\right\}+1\right)\left(\left\lfloor\frac{n}{\sqrt{L}}\right\rfloor+1\right)+j\frac{\pi}{4}\right)}, \tag{3.11}$$

where $n \in \{0, 1, \ldots, L-1\}$ and $\mathbf{c}_m$ is an m-fold wrap-around version of $\mathbf{c}_0$. By probing all the two-dimensional beamforming vectors with their specific CAZAC sequences and collecting all the receive vectors $\mathbf{y}_m$ in the matrix $\mathbf{Y} \in \mathbb{C}^{M\times L}$ as

$$\mathbf{Y} = \sqrt{P_T}\sum_{r=1}^{R}\alpha_r\mathbf{A}(\mu_r, \psi_r, \delta_p, \nu_q)\underbrace{\begin{bmatrix}\mathbf{c}_0^{\mathrm{T}}(\tau_r)\\ \mathbf{c}_1^{\mathrm{T}}(\tau_r)\\ \vdots \\ \mathbf{c}_{M-1}^{\mathrm{T}}(\tau_r)\end{bmatrix}}_{=\mathbf{C}(\tau_r)} + \underbrace{\begin{bmatrix}\mathbf{n}_0^{\mathrm{T}}\\ \mathbf{n}_1^{\mathrm{T}}\\ \vdots \\ \mathbf{n}_{M-1}^{\mathrm{T}}\end{bmatrix}}_{=\mathbf{N}}, \tag{3.12}$$

and rewriting

$$\mathbf{Y} = \sqrt{P_T}\sum_{r=1}^{R}\alpha_r\mathbf{A}(\mu_r, \psi_r, \delta_p, \nu_q)\mathbf{C}(\tau_r) + \mathbf{N}, \tag{3.13}$$

where $\mathbf{A}(\mu_r, \psi_r, \delta_p, \nu_q) \in \mathbb{C}^{M\times M}$ is represented as

$$\mathbf{A}(\mu_r, \psi_r, \delta_p, \nu_q) = \mathrm{diag}\left\{\mathbf{a}^{\mathrm{H}}(\mu_r, \psi_r)\mathbf{w}(\delta_p, \nu_q)\right\}_{m=0}^{M-1}, \tag{3.14}$$

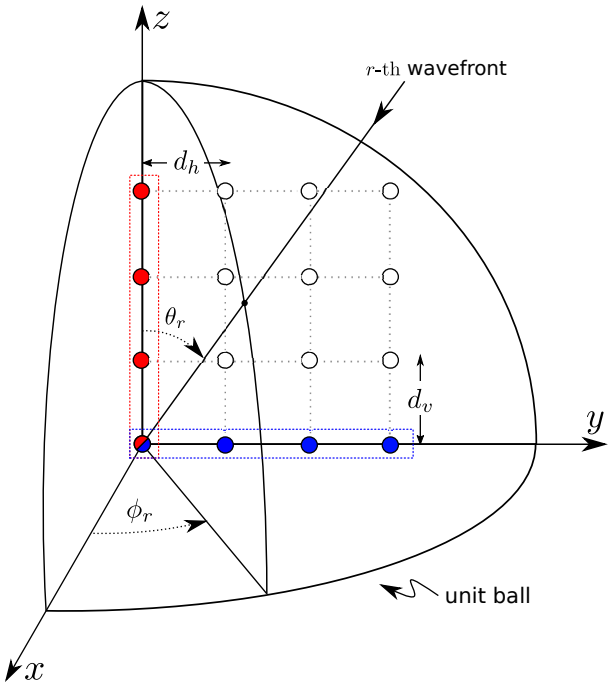


Figure 3.4 – Uniform planar array in y-z-plane.

and the noise covariance matrix is

$$\mathbf{R} = \mathrm{E}\left[\mathrm{vec}\{\mathbf{N}\}\ \mathrm{vec}\{\mathbf{N}\}^{\mathrm{H}}\right] = \sigma_n^2 \mathbf{I}_{ML}, \tag{3.15}$$

assuming spatially and temporally uncorrelated noise.

3.3.2 Coarse estimation based on modified PREIDG

In the two-stage estimation algorithm, the first stage is responsible for estimating the model order and the channel parameters which are achieved by the modified PREIDG. We use M CAZAC sequences of length $L = M$ having one-to-one correspondence to M two-dimensional beamforming vectors which are used to probe the channel. This way each UE observes M consecutive sequences $\mathbf{y}_m$ (3.5) and the received signals are multiplied with stored CAZAC sequences $\mathbf{c}_m^*$ and can be formulated for one UE as

$$\mathbf{Z} = \mathbf{Y}\mathbf{C}^{\mathrm{H}}(0) = \sqrt{P_T}\sum_{r=1}^{R} \alpha_r \mathbf{A}(\mu_r, \psi_r, \delta_p, \nu_q)\mathbf{C}(\tau_r)\mathbf{C}^{\mathrm{H}}(0) + \mathbf{N}\mathbf{C}^{\mathrm{H}}(0), \tag{3.16}$$

where $\mathbf{C}(0) \in \mathbb{C}^{M\times L}$ is the CAZAC sequences matrix with time-delay equal to zero. Further, the post-correlation power matrix $\mathbf{P}$ can be given as

$$\mathbf{P} = \mathbb{E}\left[\mathbf{Z} \odot \mathbf{Z}^*\right]. \tag{3.17}$$

Now, in any real scenario, the matrix $\mathbf{P}$ will not be sparse but can provide useful information about the model order, i.e., the number of multipath, and the channel parameters such as $\sqrt{P_T}, \alpha_r$,

ψ_r, μ_r and τ_r of each path. We exploit the power matrix by searching the (wrap-around) diagonals

$$\mathbf{p}_{i_r}^{\mathrm{T}} = \left[p_{1,i}\,,\, p_{2,\mathrm{mod}(i,M)+1}, \ldots, p_{M,\mathrm{mod}(i+M-2,M)+1}\right] \in \mathbb{R}^M, \tag{3.18}$$

where $i = 1, \ldots, M$, with $\mathbf{p}_1$ being the main diagonal of matrix $\mathbf{P}$ (3.17). For each diagonal $\mathbf{p}_i$ we check, whether the largest element is above a certain threshold G as,

$$\max_{k=1,\ldots,M} \left(p_{k,\mathrm{mod}(i+(k-2),M)+1}\right) \geq G,\ \ i = 1, \ldots, M. \tag{3.19}$$

G should be chosen such that we can detect signals above the noise floor $\sigma_n^2 M$. The integer delay τ_{i_r} for each rth path can be found as

$$\hat{\tau}_{i_r} = i_r - 1. \tag{3.20}$$

For the main diagonal $\mathbf{p}_1$, this will always be fulfilled assuming a LOS path, which has delay $\tau_i = i_1 - 1\big|_{i_1=1} = 0$. The number of diagonals $\mathbf{p}_i$, which fulfill (3.19) is the estimated model order $\hat{R}$ and provide a coarse estimate $\hat{\tau}_{i_r}$ for each path, i.e., the estimated integer delay where the index r is dropped for notational convenience in the subsequent sections.

3.3.3 LUT

To get coarse estimates of the spatial frequencies $\hat{\mu}_r$ and $\hat{\psi}_r$, we design only one LUT, which will be used for both, the azimuth beamforming angles δ_p and the elevation beamforming angles ν_q. We show this in detail for $\hat{\mu}_r$, where we use linear interpolation with $D+1$ spatial frequencies μ_d given as

$$\mu_d = \delta_p + d\Delta_\mu, \quad d = 0, \ldots, D, \tag{3.21}$$

with

$$\Delta_\mu = \frac{\delta_{p+1} - \delta_p}{D} = \frac{2\pi}{M_h D}. \tag{3.22}$$

Now, computing the hypothetical noise free normalized power using the hypothetical spatial frequencies μ_d can be given as

$$P_{p,d} = \left|\mathbf{a}^{\mathrm{H}}(\mu_d)\mathbf{w}_h(\delta_p)\right|^2, \tag{3.23}$$

and

$$P_{p+1,d} = \left|\mathbf{a}^{\mathrm{H}}(\mu_d)\mathbf{w}_h(\delta_{p+1})\right|^2, \tag{3.24}$$

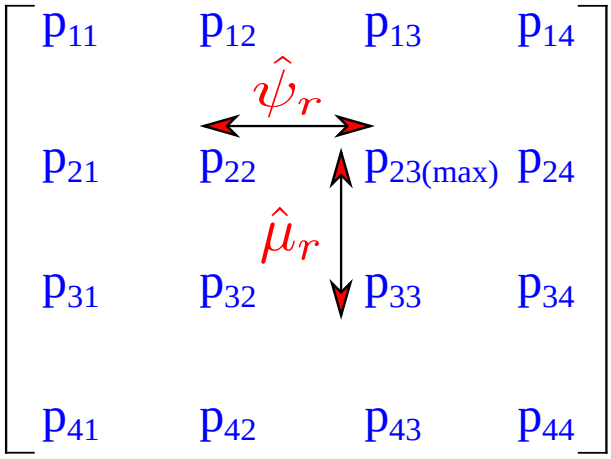


Figure 3.5 – Two-dimensional power-table $\mathbf{P}'_r$ for estimation of spatial frequencies $\hat{\mu}_r$ and $\hat{\psi}_r$.

where the channel steering vector for hypothetical spatial frequency $\mathbf{a}(\mu_d)$ can be written as

$$\mathbf{a}(\mu_d) = \left[1, e^{-j\mu_d}, \ldots, e^{-j(M_h-1)\mu_d}\right]^{\mathrm{T}} \in \mathbb{C}^{M_h \times 1}. \tag{3.25}$$

The ratio of the hypothetical powers can be given as

$$\Delta_d = \sqrt{\frac{P_{p,d}}{P_{p+1,d}}}. \tag{3.26}$$

Since the hypothetical powers $P_{p,d}$ and $P_{p+1,d}$ are independent of p, meaning, the difference between two consecutive beams is same, therefore we just need $D+1$ ratios to fill into the so-called LUT, respectively. Note that, the only one LUT, once generated is fixed and the same LUT can be used for ψ_r.

3.3.4 Estimation of Spatial Frequencies

In this section, after estimating the model order $\hat{R}$, the spatial frequencies are estimated in the following way. Let us assume $\mathbf{p}_r \in \mathbb{C}^{M \times 1}$ being the diagonal or cross diagonal vector of the received power matrix $\mathbf{P} \in \mathbb{C}^{M \times L}$ for the rth path as given in (3.18), where $M_h M_v = M$. Now, by un-vectorizing the selected $\mathbf{p}_r$ as

$$\mathbf{P}'_r = \mathrm{unvec}\{\mathbf{p}_{i_r}\}^{\mathrm{T}} \in \mathbb{C}^{M_h \times M_v}. \tag{3.27}$$

Now we have to assign to every entry of $\mathbf{P}'_r$ the beamforming vector, with which the corresponding post correlation power value has been generated.

We see that $[\mathbf{P}'_r]^{\max}_{p+1,q+1}$ as shown in Figure 3.5, i.e., the entry in the $p+1$th row and $q+1$th

Algorithm 3.1: Two-dimensional parameter estimation of $\hat{\mu}_r$ based on modified PREIDG.

1 **Require:** $\mathbf{Y}$ (3.12) ;
2 Determine $\mathbf{p}_{i_r}$ from (3.17) and (3.18);
3 Re-arrange $\mathbf{p}_{i_r}$ in the power matrix $\mathbf{P}'_r$ as shown in (3.27);
4 Find the highest power $[\mathbf{P}'_r]^{\max}_{p+1,q+1}$ and the associated beamforming vector $\mathbf{w}\left(\delta_p, \nu_q\right)$;
5 Find the second power in the $q+1$th column $[\mathbf{P}'_r]_{p+2,q+1}$ and the associated beamforming vector $\mathbf{w}\left(\delta_{p+1}, \nu_q\right)$;
6 Calculate Δ_{μ_r} as in (3.29);
7 Find d such that $\Delta_d \geq \Delta_{\mu_r} \geq \Delta_{d+1}$ as in (3.26) ;
8 Calculate constant b_μ as in (3.31);
9 **Return** $\hat{\mu}_r$ as in (3.30). ;

column of $\mathbf{P}'_r$ has been generated by the following beamforming vector,

$$\mathbf{w}\left(\delta_p, \nu_q\right) = \mathbf{w}_h(\delta_p) \otimes \mathbf{w}_v(\nu_q). \tag{3.28}$$

Let us assume that the second largest entry in the $p+1$th row is $[\mathbf{P}'_r]_{p+1,q}$ and in the $q+1$th column is $[\mathbf{P}'_r]_{p+2,q+1}$. Then the interpolation for μ_r should be carried out between $\delta_p < \hat{\mu}_r < \delta_{p+1}$ and for ψ_r between $\nu_{q-1} < \hat{\psi}_r < \nu_q$.

Now let us use the LUT which we produced according to (3.21)-(3.26) and let us choose those indices d and $d+1$ such that Δ_d and Δ_{d+1} are the two ratios closest to

$$\Delta_{\mu_r} = \sqrt{\frac{[\mathbf{P}'_r]_{p+1,q+1}}{[\mathbf{P}'_r]_{p+2,q+1}}}. \tag{3.29}$$

We estimate μ_r as

$$\hat{\mu}_r = \mu_d + b_\mu \Delta_\mu, \tag{3.30}$$

with

$$b_\mu = \frac{\Delta_d - \Delta_{\mu_r}}{\Delta_d - \Delta_{d+1}}. \tag{3.31}$$

The μ_r estimation approach is given in Algorithm 3.1. The same procedure can be adapted for estimating $\hat{\psi}_r$.

3.3.5 Example with a 4×4 *URA*

Let us take 4×4 URA having $M_h = 4$ and $M_v = 4$ with the total number of antennas at each subarray $M = 16$. Therefore, by defining the horizontal and vertical beamforming vectors as

$$\mathbf{w}_h(\delta_p) = \frac{1}{2}\left[1, e^{-j\delta_p}, e^{-j2\delta_p}, e^{-j3\delta_p}\right]^{\mathrm{T}}, \tag{3.32}$$

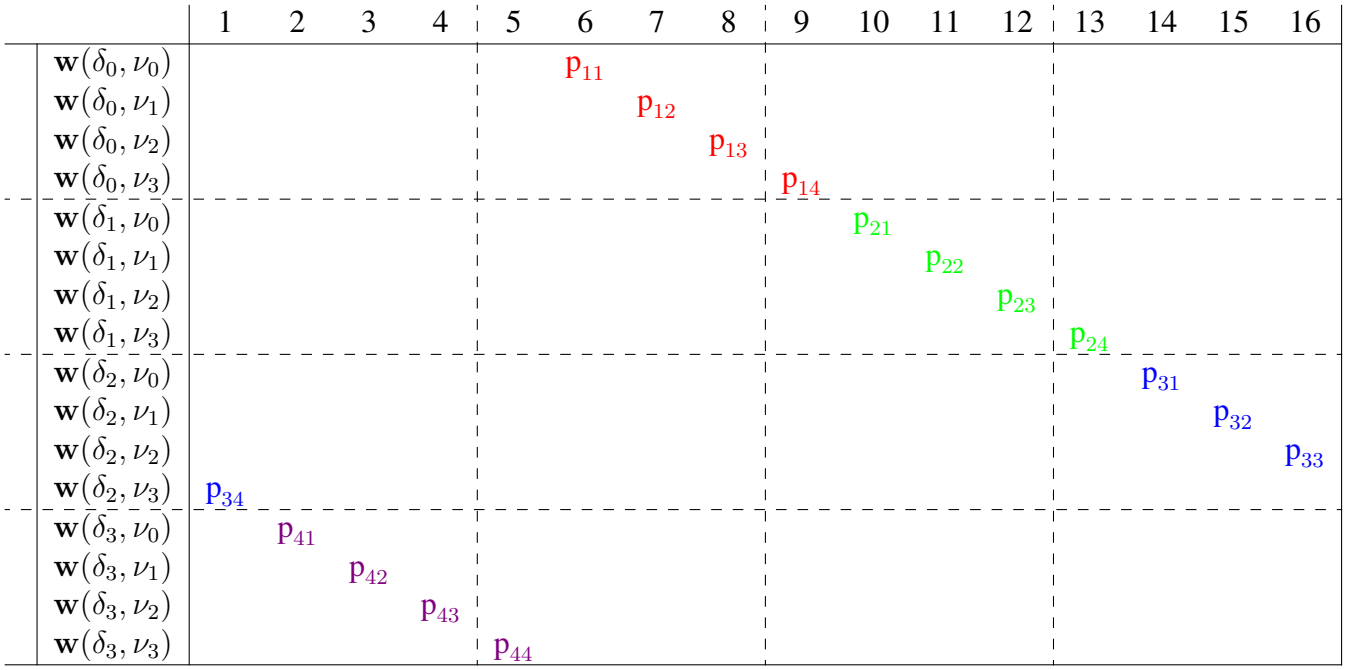


Figure 3.6 – An example for constructing received power matrix $\mathbf{P}$ considering a 4×4 URA.

and

$$\mathbf{w}_v(\nu_q) = \frac{1}{2}\left[1, e^{-j\nu_q}, e^{-j2\nu_q}, e^{-j3\nu_q}\right]^{\mathrm{T}}, \tag{3.33}$$

where $\delta_0 = \nu_0 = 0$, $\delta_1 = \nu_1 = \frac{\pi}{2}$, $\delta_2 = \nu_2 = \pi$, $\delta_3 = \nu_3 = \frac{3\pi}{2}$. The resultant two-dimensional beamforming vector can be written as

$$\begin{aligned}&\mathbf{w}(\delta_p,\nu_q) = \mathbf{w}(\delta_p) \otimes \mathbf{w}(\nu_q) = \\ &\frac{1}{4}\left[1, e^{-j\nu_q}, e^{-j2\nu_q}, e^{-j3\nu_q}, e^{-j\delta_p}, e^{-j(\delta_p+\nu_q)}, e^{-j(\delta_p+2\nu_q)}\right. \\ &\left. e^{-j(\delta_p+3\nu_q)}, \ldots, e^{-j(3\delta_p+2\nu_q)}, e^{-j(3\delta_p+3\nu_q)}\right]^{\mathrm{T}}.\end{aligned} \tag{3.34}$$

When generating the power matrix $\mathbf{P}$ (3.17), each row is obtained with one beamforming vector $\mathbf{w}(\delta_p,\nu_q)$, i.e., with one index pair (p,q) as shown in Figure 3.6.

Let us assume that the rth path has a integer-delay $\tau_{i_r} = 6$. The diagonal $\mathbf{p}_{i_r}$ looks as follows:

$$\mathbf{p}_{i_r} = [p_{11}, \ldots, p_{14}, p_{21}, \ldots, p_{24}, p_{31}, \ldots, p_{34}, p_{41}, \ldots, p_{44}]^{\mathrm{T}}, \tag{3.35}$$

and

$$\mathbf{P}'_r = \mathrm{unvec}\{\mathbf{p}_{i_r}\}^{\mathrm{T}} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix}. \tag{3.36}$$

Therefore, the power entry $[\mathbf{P}'_r]_{p+1,q+1}$ has been produced with $\mathbf{w}(\delta_p, \nu_q)$.
Now let us take an example for a demonstration with a UE position at AoD azimuth $\phi = 35°$ and elevation $\theta = 95°$ as shown in Figure 3.7. The largest element of $\mathbf{P}'_r$ is p_{21}, the two next largest ones are p_{31} and p_{24}. The three corresponding beamforming vectors should be

$$\begin{aligned}
&\text{Beam 1} : \mathbf{w}(\delta_1, \nu_0) \;\text{ with }\; \phi_{1,0} = 30°, \theta_{1,0} = 90° \\
&\text{Beam 2} : \mathbf{w}(\delta_2, \nu_0) \;\text{ with }\; \phi_{2,0} = 90°, \theta_{2,0} = 90° \\
&\text{Beam 3} : \mathbf{w}(\delta_1, \nu_3) \;\text{ with }\; \phi_{1,3} = 35.2644°, \theta_{1,3} = 120°.
\end{aligned}$$

The corresponding positions in the $\mathbf{P}'_r$ matrix should be for

$$\begin{aligned}
&\text{Beam 1} : \;\; [\mathbf{P}'_r]_{2,1} = [\mathbf{P}'_r]^{\max}_{2,1} \\
&\text{Beam 2} : [\mathbf{P}'_r]_{3,1} \\
&\text{Beam 3} : [\mathbf{P}'_r]_{2,4}\,.
\end{aligned}$$

Therefore, the interpolation for $\hat{\mu}_r$ should take place between δ_1 and δ_2 which correspond to beam 1 and beam 2 respectively. Similarly, the interpolation for $\hat{\psi}_r$ should take place between ν_3 and ν_0, which correspond beam 3 and beam 1 respectively.

Since δ_p and ν_q in the DFT matrices are in the range $[0, 2\pi)$, and Algorithm 3.1 considers them in the range $(-\pi, \pi]$, the following transformation is applied:

$$x = \mathrm{mod}(x + \pi, 2\pi) - \pi, \quad x \in \{\delta_p, \nu_q, \hat{\mu}_r, \hat{\psi}_r\}. \tag{3.37}$$

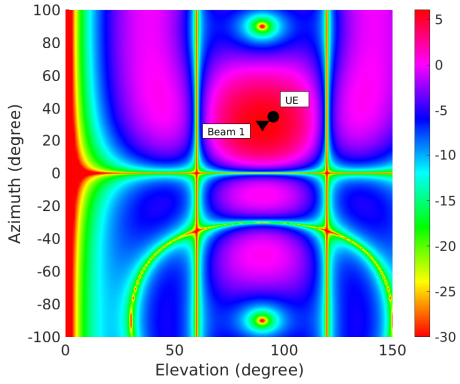
Finally, to get an estimate for the azimuth $\hat{\phi}_r$ and the elevation $\hat{\theta}_r$ as,

$$\hat{\theta}_r = \arccos\left(\frac{\hat{\psi}_r}{\pi}\right), \tag{3.38}$$

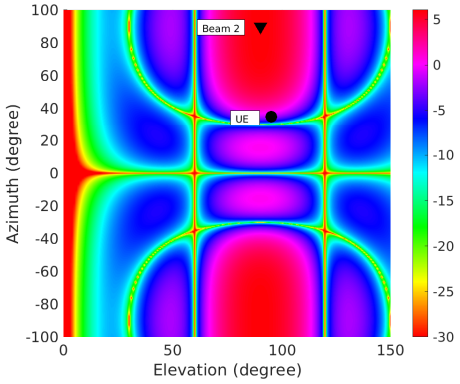
$$\hat{\phi}_r = \arcsin\left(\frac{\hat{\mu}_r}{\pi \sin(\hat{\theta}_r)}\right), \tag{3.39}$$

with the range of $\theta_r^\circ \sim \mathrm{U}(90°, 130°)$ and $\phi_r^\circ \sim \mathrm{U}(-60°, 60°)$.

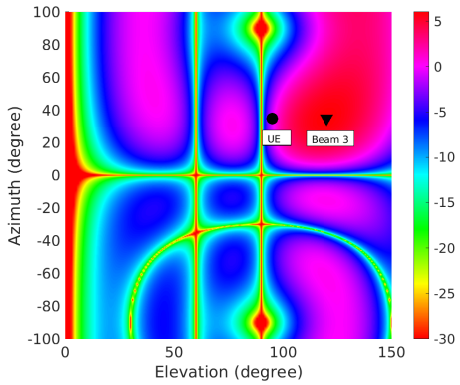
Note: The interpolation between δ_p and δ_{p+1} for estimating μ_r, and similarly the interpolation between ν_{q-1} and ν_q for estimating ψ_r can go wrong in some cases, especially if the signal level is weak. Consequently, let us say that μ_r is close to the δ_p with highest received power, but the received power generated using δ_{p+1} may be close to the noise floor and similarly let us say that



(a) Maximum power received using beam-1 $\mathbf{w}\,(\phi_{1,0} = 30°, \theta_{1,0} = 90°)$



(b) Second-highest power received using beam-2 $\mathbf{w}\,(\phi_{2,0} = 90°, \theta_{2,0} = 90°)$



(c) Second-highest power received using beam-3 $\mathbf{w}\,(\phi_{1,3} = 35.2644°, \theta_{1,3} = 120°)$

Figure 3.7 – An example of two-dimensional coarse estimation via three two-dimensional beamforming vectors for the UE at $(\phi = 35°, \theta = 95°)$

ψ_r is close to ν_q with highest received power but the received power generated using ν_{q-1} may be close to the noise floor. Therefore, we check whether

$$\begin{cases} |[\mathbf{P}'_r]_{p+1,q+2} - [\mathbf{P}'_r]_{p+1,q}| \leq \sigma_n^2, \\ |[\mathbf{P}'_r]_{p+2,q+1} - [\mathbf{P}'_r]_{p,q+1}| \leq \sigma_n^2. \end{cases} \tag{3.40}$$

If (3.40) is fulfilled, then it is not worthwhile to interpolate at all, but simply choose $\hat{\mu}_r = \delta_p$, and $\hat{\psi}_r = \nu_q$.

3.3.6 High-resolution channel parameter estimation using SAGE

High-resolution parameter estimation in general is of practical interest for wireless communication systems. High-resolution channel parameter estimation plays an important role in designing hybrid precoders at the BS because the channel parameters need to be quantized before being sent back to the BS. The BS receives the channel parameters with quantization error resulting in deterioration of the performance. The deterioration can be increased if the channel

parameters are not estimated with high accuracy.

It is also known in the literature that SAGE and ESPRIT are high-resolution estimation algorithms [36], but the SAGE algorithm can provide the highest accuracy as compared to MUSIC and ESPRIT [37, 56, 57, 58]. On the the other hand, ESPRIT has less computational complexity as compared to the standard SAGE algorithm, however, ESPRIT needs two translation invariant subarrays in the URA while on the other hand SAGE does not impose such kind of an assumption. To ensure a high-resolution parameter estimation with less computational complexity, we initialize the SAGE algorithm with coarse estimates achieved by the modified PREIDG algorithm. We use the fundamental principles of the SAGE algorithm [59, 49] to derive new expressions for our non-linear problem at hand. Furthermore, using SAGE, a refined estimate of complex path gain and non-integer delay can be achieved.

Let us assume a random variable $\mathbf{Y}$ as given in (3.12) having a multivariate complex Gaussian pdf which is parametrized by the unknown channel parameter vector $\boldsymbol{\eta}$ as

$$\boldsymbol{\eta} = \left[\sqrt{P_T}\mathrm{Re}\{\boldsymbol{\alpha}\}^{\mathrm{T}}, \sqrt{P_T}\mathrm{Im}\{\boldsymbol{\alpha}\}^{\mathrm{T}}, \boldsymbol{\mu}^{\mathrm{T}}, \boldsymbol{\psi}^{\mathrm{T}}, \boldsymbol{\tau}^{\mathrm{T}}\right]^{\mathrm{T}}, \tag{3.41}$$

where

$$\sqrt{P_T}\mathrm{Re}\{\boldsymbol{\alpha}\} = \left[\sqrt{P_T}\mathrm{Re}\{\alpha_1\}, \ldots, \sqrt{P_T}\mathrm{Re}\{\alpha_R\}\right]^{\mathrm{T}} \in \mathbb{C}^{R\times 1}, \tag{3.42}$$

$$\sqrt{P_T}\mathrm{Im}\{\boldsymbol{\alpha}\} = \left[\sqrt{P_T}\mathrm{Im}\{\alpha_1\}, \ldots, \sqrt{P_T}\mathrm{Im}\{\alpha_R\}\right]^{\mathrm{T}} \in \mathbb{C}^{R\times 1}, \tag{3.43}$$

$$\boldsymbol{\mu} = [\mu_1, \ldots, \mu_R]^{\mathrm{T}} \in \mathbb{C}^{R\times 1}, \tag{3.44}$$

$$\boldsymbol{\psi} = [\psi_1, \ldots, \psi_R]^{\mathrm{T}} \in \mathbb{C}^{R\times 1}, \tag{3.45}$$

$$\boldsymbol{\tau} = [\tau_1, \ldots, \tau_R]^{\mathrm{T}} \in \mathbb{C}^{R\times 1}. \tag{3.46}$$

Thus, the likelihood can be given as

$$\mathrm{L}(\mathbf{Y};\boldsymbol{\eta}) = \frac{1}{\pi^{ML}\det\mathbf{R}}\ \exp\left(-\mathrm{vec}\left\{\mathbf{Y} - \sqrt{P_T}\sum_{r=1}^{R}\alpha_r\mathbf{A}(\mu_r,\psi_r,\delta_p,\nu_q)\mathbf{C}(\tau_r)\right\}^{\mathrm{H}}\right.$$
$$\left.\mathbf{R}^{-1}\ \mathrm{vec}\left\{\mathbf{Y} - \sqrt{P_T}\sum_{r=1}^{R}\alpha_r\mathbf{A}(\mu_r,\psi_r,\delta_p,\nu_q)\mathbf{C}(\tau_r)\right\}\right). \tag{3.47}$$

The ML estimator is given as

$$\hat{\boldsymbol{\eta}} = \arg\max_{\boldsymbol{\eta}} \mathrm{L}\left(\mathbf{Y};\boldsymbol{\eta}\right) \tag{3.48}$$

whereas no closed form solution can found for the ML estimator as mention in (3.48) because it is a multidimensional, non-linear optimization problem. We use SAGE algorithm [49] to

solve our non-linear problem considering important hardware constraints. We configure the SAGE algorithm for our two-dimensional channel scenario and considering one-dimensional optimization problems with respect to the sequence of parameter estimates. For notational convenience, we represent $\mathbf{A}(\mu_r, \psi_r, \delta_p, \nu_q)$ as $\mathbf{A}(\mu_r, \psi_r)$.

Thus, the parameters of the rth path can be estimated with the following expectation-maximization steps

Expectation step:

$$\hat{\mathbf{X}}_r = \mathbf{Y} - \sum_{\substack{r'=1 \\ r' \neq r}}^{R} \mathbf{S}_{r'}(\hat{\boldsymbol{\eta}}_{r'}), \tag{3.49}$$

where

$$\mathbf{S}_r(\boldsymbol{\eta}_r) = \sqrt{P_T}\alpha_r \mathbf{A}(\mu_r, \psi_r)\mathbf{C}(\tau_r). \tag{3.50}$$

Maximization step:

$$\hat{\tau}_r = \arg\max_{\tau_r} \left\{ \frac{\left| \mathrm{tr}\left\{ \mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\hat{\mu}_r, \hat{\psi}_r)\hat{\mathbf{X}}_r \right\} \right|^2}{\beta_r \sigma_n^2 \mathrm{tr}\left\{ \mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\hat{\mu}_r, \hat{\psi}_r)\mathbf{A}(\hat{\mu}_r, \hat{\psi}_r)\mathbf{C}(\tau_r) \right\}} \right\}, \tag{3.51}$$

the spatial frequency $\hat{\mu}_r$ can be iteratively estimated as

$$\hat{\mu}_r = \arg\max_{\mu_r} \left\{ \frac{\left| \mathrm{tr}\left\{ \mathbf{C}^{\mathrm{H}}(\hat{\tau}_r)\mathbf{A}^{\mathrm{H}}(\mu_r, \hat{\psi}_r)\hat{\mathbf{X}}_r \right\} \right|^2}{\beta_r \sigma_n^2 \mathrm{tr}\left\{ \mathbf{C}^{\mathrm{H}}(\hat{\tau}_r)\mathbf{A}^{\mathrm{H}}(\mu_r, \hat{\psi}_r)\mathbf{A}(\mu_r, \hat{\psi}_r)\mathbf{C}(\hat{\tau}_r) \right\}} \right\}, \tag{3.52}$$

and the spatial frequency $\hat{\psi}_r$ can be iteratively estimated as

$$\hat{\psi}_r = \arg\max_{\psi_r} \left\{ \frac{\left| \mathrm{tr}\left\{ \mathbf{C}^{\mathrm{H}}(\hat{\tau}_r)\mathbf{A}^{\mathrm{H}}(\hat{\mu}_r, \psi_r)\hat{\mathbf{X}}_r \right\} \right|^2}{\beta_r \sigma_n^2 \mathrm{tr}\left\{ \mathbf{C}^{\mathrm{H}}(\hat{\tau}_r)\mathbf{A}^{\mathrm{H}}(\hat{\mu}_r, \psi_r)\mathbf{A}(\hat{\mu}_r, \psi_r)\mathbf{C}(\hat{\tau}_r) \right\}} \right\}. \tag{3.53}$$

Finally, the complex path gain $\sqrt{\hat{P_T}}\alpha_r$ can be analytically found as

$$\sqrt{\hat{P_T}}\alpha_r = \frac{\mathrm{tr}\left\{ \mathbf{C}^{\mathrm{H}}(\hat{\tau}_r)\mathbf{A}^{\mathrm{H}}(\hat{\mu}_r, \hat{\psi}_r)\hat{\mathbf{X}}_r \right\}}{\mathrm{tr}\{\mathbf{C}^{\mathrm{H}}(\hat{\tau}_r)\mathbf{A}^{\mathrm{H}}(\hat{\mu}_r, \hat{\psi}_r)\mathbf{A}(\hat{\mu}_r, \hat{\psi}_r)\mathbf{C}(\hat{\tau}_r)\}}. \tag{3.54}$$

The derivations are given in Appendix B.1.

3.3.6.1 Convergence analysis of SAGE

The second stage of the high-resolution algorithm is to refine the coarse estimates achieved via modified PREIDG by the SAGE algorithm. To make sure, that the SAGE algorithm

converges to the global optimum using very few iterations is of practical interest. For this reason, it is mandatory to initialize the SAGE algorithm with estimates that are already close to the global optimum to guarantee convergence to the global optimum in only a few iterations. We use $\hat{\tau}_i$ (3.20), $\hat{\mu}_r$ (3.30), $\hat{\psi}_r$, and $\hat{\alpha}_r = 0$ to initialize the SAGE algorithm. One iteration of the SAGE algorithm is a full update of the parameter vector $\boldsymbol{\eta}$. The stopping thresholds for the SAGE algorithm's convergence are given as

$$\mathrm{T}_1 = \frac{|\hat{\psi}_{r_p} - \hat{\psi}_r|}{|\hat{\psi}_r|}, \tag{3.55}$$

$$\mathrm{T}_2 = \frac{|\hat{\mu}_{r_p} - \hat{\mu}_r|}{|\hat{\mu}_r|}, \tag{3.56}$$

$$\mathrm{T}_3 = \frac{|\hat{\tau}_{r_p} - \hat{\tau}_r|}{|\hat{\tau}_r|}, \tag{3.57}$$

$$\mathrm{T}_4 = \frac{|\sqrt{\hat{P}_T}\alpha_{r_p} - \sqrt{\hat{P}_T}\alpha_r|}{|\sqrt{\hat{P}_T}\alpha_r|}, \tag{3.58}$$

where $\hat{\psi}_{r_p}, \hat{\mu}_{r_p}, \hat{\tau}_{r_p}, \sqrt{\hat{P}_T}\alpha_{r_p}$, are the previous estimates of spatial frequencies, time-delay, and complex path gain. Convergence of the SAGE algorithm is achieved if

$$\max\{\mathrm{T}_1, \mathrm{T}_2, \mathrm{T}_3, \mathrm{T}_4\} \leq \Gamma, \tag{3.59}$$

where Γ represents the stopping threshold. The performance of the SAGE algorithm can be improved by reducing the stopping threshold at the cost of more iterations.

One must note that, if the model order is estimated wrongly at the coarse estimation stage and feed the same model order to SAGE, it is highly probable that the SAGE algorithm can ends up with an outlier in the refinement stage.

3.3.7 Complexity of the proposed two-step approach

Recall that the proposed algorithm has two steps. In the first step, the modified PREIDG is proposed, which is the received signal matrix $\mathbf{Y}$, filtered out using the stored CAZAC sequence matrix $\mathbf{C}(0)$, as shown in (3.16), the complexity of which corresponds to that of the matrix product $\mathcal{O}\left(4M^2L\right)$. As far as the second step is concerned, which is the SAGE algorithm, therefore the complexity involves the computation of $\hat{\tau}_r$ (3.51), $\hat{\mu}_r$ (3.52), $\hat{\psi}_r$ (3.53), and, $\hat{\alpha}_r$ (3.54). Summing up the number of operations defined by these equations, and assuming J iterations for the convergence, we arrive at $\mathcal{O}\left(4J\left(12M^2L + 8L^2M\right)\right)$. The overall complexity is therefore $\mathcal{O}\left(4J\left(12M^2L + 8L^2M\right) + 4M^2L\right)$.

3.3.8 CRLB

In this section, we derive the CRLB for the spatial frequencies ψ_r, μ_r, complex path-gain $\sqrt{P_T}\alpha_r$ and the time-delay τ_r for each path, which is given in (3.41).
Assuming $\hat{\boldsymbol{\eta}}$ as an unbiased estimate of $\boldsymbol{\eta}$, then the variance of the estimation error can be lower-bounded by the diagonal elements of the inverse of the FIM, $\mathbf{F}(\boldsymbol{\eta})$ [46]

$$\mathrm{var}(\hat{\eta}_i) \geq \left[\mathbf{F}^{-1}(\boldsymbol{\eta})\right]_{i\,i}. \tag{3.60}$$

The lower bound of the standard deviation of the estimation error can be given as

$$\sqrt{CRLB(\hat{\eta}_i)} = \sqrt{[\mathbf{F^{-1}}(\boldsymbol{\eta})]_{ii}}. \tag{3.61}$$

The FIM for the complex data can be given as [46]

$$[\mathbf{F}(\boldsymbol{\eta})]_{ij} = \frac{2}{\sigma_n^2}\ \mathrm{Re}\ \left\{\mathrm{tr}\left\{\frac{\partial \mathbf{S}^{\mathrm{H}}(\boldsymbol{\eta})}{\partial \eta_i}\ \frac{\partial \mathbf{S}(\boldsymbol{\eta})}{\partial \eta_j}\right\}\right\}, \tag{3.62}$$

where the matrix $\mathbf{S}(\boldsymbol{\eta})$

$$\mathbf{S}(\boldsymbol{\eta}) = \sum_{r=1}^{R} \mathbf{S}\left(\boldsymbol{\eta}_r\right) = \sqrt{P_T}\ \sum_{r=1}^{R} \alpha_r\, \mathbf{A}(\mu_r, \psi_r)\, \mathbf{C}(\tau_r). \tag{3.63}$$

The FIM $\mathbf{F}(\boldsymbol{\eta})$ can be structured as

$$\mathbf{F}(\boldsymbol{\eta}) = \begin{bmatrix} \mathbf{F}_{\mathrm{Re}\{\boldsymbol{\alpha}\}\mathrm{Re}\{\boldsymbol{\alpha}\}} & \mathbf{F}_{\mathrm{Re}\{\boldsymbol{\alpha}\}\mathrm{Im}\{\boldsymbol{\alpha}\}} & \mathbf{F}_{\mathrm{Re}\{\boldsymbol{\alpha}\}\boldsymbol{\mu}} & \mathbf{F}_{\mathrm{Re}\{\boldsymbol{\alpha}\}\boldsymbol{\psi}} & \mathbf{F}_{\mathrm{Re}\{\boldsymbol{\alpha}\}\boldsymbol{\tau}} \\ \mathbf{F}^{\mathrm{T}}_{\mathrm{Re}\{\boldsymbol{\alpha}\}\mathrm{Im}\{\boldsymbol{\alpha}\}} & \mathbf{F}_{\mathrm{Im}\{\boldsymbol{\alpha}\}\mathrm{Im}\{\boldsymbol{\alpha}\}} & \mathbf{F}_{\mathrm{Im}\{\boldsymbol{\alpha}\}\boldsymbol{\mu}} & \mathbf{F}_{\mathrm{Im}\{\boldsymbol{\alpha}\}\boldsymbol{\psi}} & \mathbf{F}_{\mathrm{Im}\{\boldsymbol{\alpha}\}\boldsymbol{\tau}} \\ \mathbf{F}^{\mathrm{T}}_{\mathrm{Re}\{\boldsymbol{\alpha}\}\boldsymbol{\mu}} & \mathbf{F}^{\mathrm{T}}_{\mathrm{Im}\{\boldsymbol{\alpha}\}\boldsymbol{\mu}} & \mathbf{F}_{\boldsymbol{\mu}\boldsymbol{\mu}} & \mathbf{F}_{\boldsymbol{\mu}\boldsymbol{\psi}} & \mathbf{F}_{\boldsymbol{\mu}\boldsymbol{\tau}} \\ \mathbf{F}^{\mathrm{T}}_{\mathrm{Re}\{\boldsymbol{\alpha}\}\boldsymbol{\psi}} & \mathbf{F}^{\mathrm{T}}_{\mathrm{Im}\{\boldsymbol{\alpha}\}\boldsymbol{\psi}} & \mathbf{F}^{\mathrm{T}}_{\boldsymbol{\mu}\boldsymbol{\psi}} & \mathbf{F}_{\boldsymbol{\psi}\boldsymbol{\psi}} & \mathbf{F}_{\boldsymbol{\psi}\boldsymbol{\tau}} \\ \mathbf{F}^{\mathrm{T}}_{\mathrm{Re}\{\boldsymbol{\alpha}\}\boldsymbol{\tau}} & \mathbf{F}^{\mathrm{T}}_{\mathrm{Im}\{\boldsymbol{\alpha}\}\boldsymbol{\tau}} & \mathbf{F}^{\mathrm{T}}_{\boldsymbol{\mu}\boldsymbol{\tau}} & \mathbf{F}^{\mathrm{T}}_{\boldsymbol{\psi}\boldsymbol{\tau}} & \mathbf{F}_{\boldsymbol{\tau}\boldsymbol{\tau}} \end{bmatrix}. \tag{3.64}$$

The blocks of $\mathbf{F}(\boldsymbol{\eta})$ are derived in Appendix B.2.

3.3.9 Numerical Results

In this section, the performance of the proposed modified PREIDG is compared to the state-of-the-art ABP algorithm [27] and the high-resolution parameter estimation achieved via SAGE is compared to the CRLB, respectively. The transmitter at the BS is equipped with a PCBMS architecture for exciting the URA having $M = M_h M_v = 16$ assuming the inter-element spacing $d_h = d_v = \lambda/2$. We use one subarray of size 4×4 for parameter estimation assuming round-robin fashion. We assume the carrier frequency $f_c = 28\,\mathrm{GHz}$ and bandwidth of the system

as $B = 200\,\mathrm{MHz}$ giving one symbol duration $T_s = 5\,\mathrm{ns}$. We use 80 pilot symbols per beam for channel probing which leads to the total number of 800 QPSK symbols as pilot overhead. We model the channel parameters based on [47]. The total number of paths is assumed as $R = 3$. The LOS distance between the BS and UE is uniformly distributed as

$$X_{\text{los}} \sim \mathrm{U}(30\,\mathrm{m}, 60\,\mathrm{m}), \tag{3.65}$$

similarly, the NLOS distances are uniformly distributed as

$$X_{\text{nlos}} = X_{\text{los}} + \Delta_{\text{nlos}}, \tag{3.66}$$

where Δ_{nlos} is the relative distance between the LOS and NLOS path, distributed as $\Delta_{\text{nlos}} \sim \mathrm{U}(4.5\,\mathrm{m}, 24\,\mathrm{m})$ which ends up with the delay difference between 3 and 16 symbols. The length of the CAZAC sequence $L = 16$. The length of the CAZAC sequence determines the maximum time-delay of the NLOS paths to be estimated. The path loss can be given as

$$PL(\mathrm{dB}) = 10\bar{n}\log_{10}\left(\frac{X}{X_0}\right), \tag{3.67}$$

where $\bar{n}$ represents the path loss exponent which is chosen as 2.1 and 2.4 for LOS and NLOS, respectively. X_0 is assumed as $1\,\mathrm{m}$. γ_r is the magnitude of the complex path gain of each path-r and is defined as the ratio of the path loss for LOS (PL_{los}) and NLOS (PL_{nlos}) respectively. Note: we are interested in the ratio of LOS and NLOS path

$$\gamma_r = \sqrt{\frac{PL_{\text{los}}}{PL_{\text{nlos}}}}, \tag{3.68}$$

The complex path gain α_r for each path can be given as

$$\frac{\alpha_r}{\alpha_1} = \gamma_r e^{j\varphi_r}, \tag{3.69}$$

where α_1 is the complex path gain for the LOS signal and is assumed as one because only the ratio of the LOS path and NLOS paths does matter in a real scenario. φ_r is the phase of the complex path coefficient for each path and is generated as $\varphi_r \sim \mathrm{U}(0, 2\pi)$. Note that, the phase of the complex gain φ_r is different for each sub-array at the BS while the magnitude γ_r is the same for each sub-array. In this approach, as the BS knows the geometry of the URA, the BS can estimate the phase of the other sub-arrays based on the estimation of the phase $\hat{\varphi}_1$ of the first sub-array. The AoDs azimuth ϕ_r and elevation θ_r for LOS and NLOS can be generated

assuming one sector of a cell as $\phi_r^\circ \sim \mathrm{U}(-60^\circ, +60^\circ)$ and $\theta_r^\circ \sim \mathrm{U}(90^\circ, 130^\circ)$, respectively. The noise variance is assumed as $\sigma_n^2 = 1$. The SNR is defined as

$$\text{SNR} = \frac{P_T|\alpha_1|^2}{\sigma_n^2}. \tag{3.70}$$

The RMSE for the different channel parameters can be define as

$$\text{RMSE}(\hat{\phi}_r) = \sqrt{\mathbb{E}|\phi_r - \hat{\phi}_r|^2}, \tag{3.71}$$

$$\text{RMSE}(\hat{\theta}_r) = \sqrt{\mathbb{E}|\theta_r - \hat{\theta}_r|^2}, \tag{3.72}$$

$$\text{RMSE}(\hat{\sqrt{P_T}\alpha_r}) = \sqrt{\mathbb{E}\left|\frac{\sqrt{P_T}\alpha_r - \hat{\sqrt{P_T}\alpha_r}}{\sqrt{P_T}\alpha_r}\right|^2}, \tag{3.73}$$

$$\text{RMSE}(\hat{\tau}_r) = \sqrt{\mathbb{E}|\tau_r - \hat{\tau}_r|^2}. \tag{3.74}$$

The CRLB on the error is calculated for assessment as

$$\sqrt{\text{CRLB}(\hat{\eta}_i)} = \sqrt{[\mathbf{F}^{-1}(\boldsymbol{\eta})]_{i\,i}}. \tag{3.75}$$

We use 10000 channel realizations for calculating the RMSE for each parameter at different SNR. Similarly, the CRLB, each parameter is obtained by numerical averaging over 10000 channel scenarios. The length of the LUT for the modified PREIDG algorithm is $D = 1001$.

To form beam pairs in the ABP method, we fixed $\delta = \frac{2m\pi}{M_h} = \frac{\pi}{2}$, assuming $m = 1$ as explained in [27]. There are a total of 8 beam-pairs that can be formed for the estimation of ψ_r and μ_r. For the ABP method, *a priori* knowledge has been assumed about the auxiliary beam-pair used for the estimation of spatial frequencies.

The motivation behind the modified PREIDG based coarse estimation is to initialize the SAGE algorithm for getting the high-resolution parameter estimates with a reduced number of iterations for convergence, simply means to achieve the global optimum. With the initialization of the SAGE algorithm with $\hat{\psi}_r$, $\hat{\mu}_r$, $\hat{\tau}_i$ and assuming $\sqrt{P_T}\alpha_r = 0$ with fixing the threshold $\Gamma = 10^{-3}$, facilitate the convergence of the SAGE algorithm with maximum number of 6 iterations but in 70% of the channel realizations, SAGE took 4 iterations to converge, which is of practical interest as shown in Figure 3.8.

Figure 3.9 shows the performance of the proposed modified PREIDG algorithm as compared with the ABP method for the AoDs azimuth ϕ_r LOS only. The novel modified PREIDG based parameter estimation marked better performance as compared to the ABP method. The ABP method can not show an improvement because of the limited beam pairs formed/available. The SAGE algorithm outperforms the ABP method because of its high-resolution capability. The

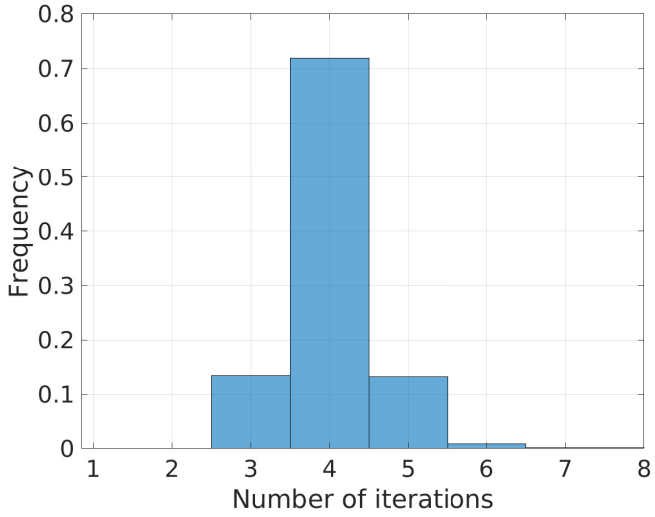


Figure 3.8 – Histogram for the number of iterations of SAGE in the two-stage algorithm.

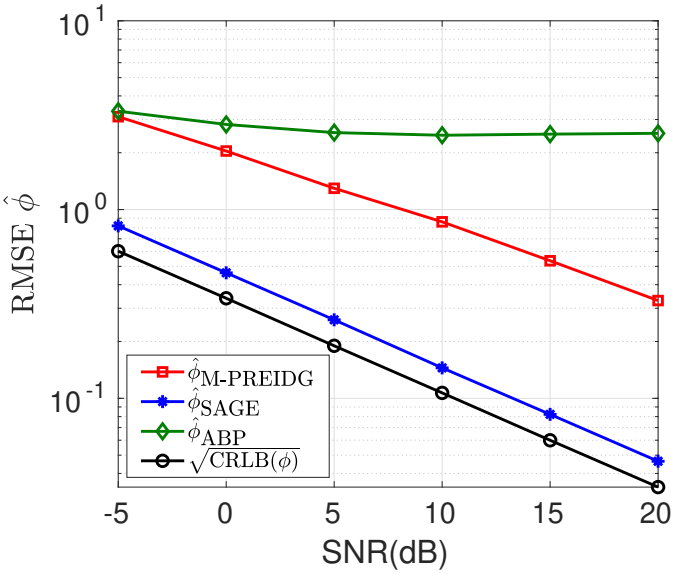


Figure 3.9 – Performance comparison of two-stage and ABP algorithm for LOS AoD azimuth, assuming $R = 3$.

SAGE algorithm close to the CRLB and thus proofs to be efficient.

Similarly, the novel modified PREIDG algorithm shows good estimation performance as compared to the ABP method for AoDs azimuth ϕ_r NLOS paths only. The SAGE algorithm shows a better performance with less number of iterations as shown in Figure 3.10. Furthermore, the SAGE algorithm approach close to the CRLB. The ABP method can not improve the performance because of the limited beam pairs probed for estimation. The performance of our novel modified PREIDG is evaluated against ABP method for AoD elevation θ_r for LOS only as shown in Figure 3.11. The modified PREIDG based estimation shows good performance as compared to the ABP method. The SAGE estimation throughout outperforms and approach close to the

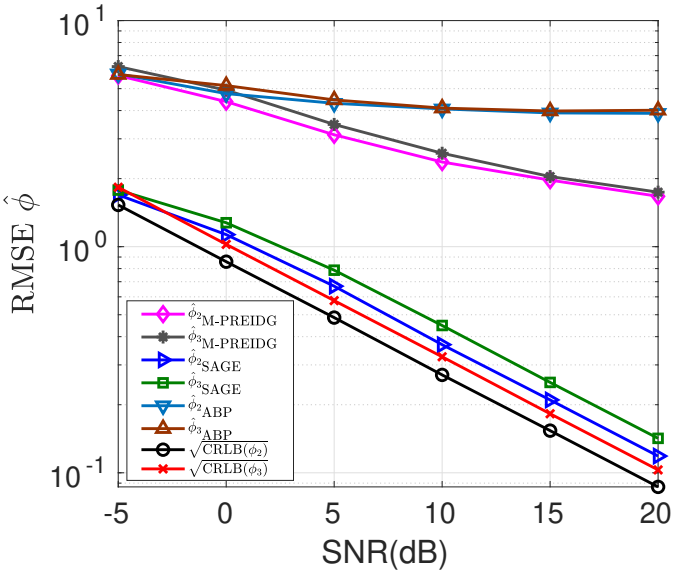


Figure 3.10 – Performance comparison of two-stage and ABP algorithm for NLOS AoDs azimuth, assuming $R = 3$.

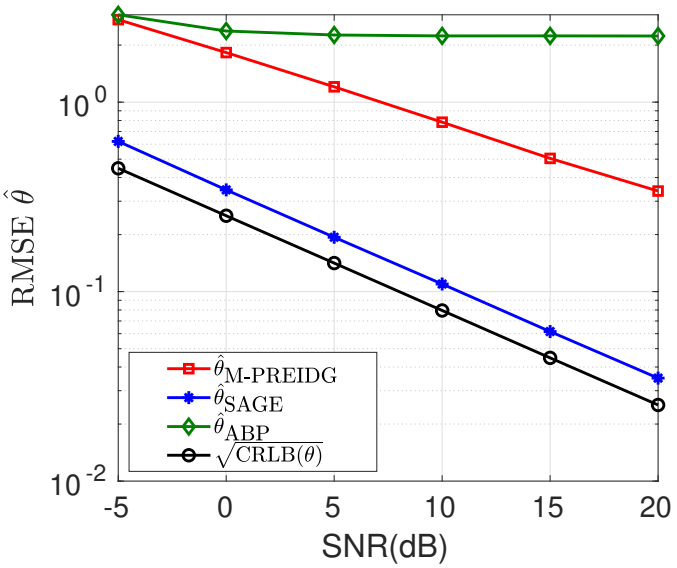


Figure 3.11 – Performance comparison of two-stage and ABP algorithm for LOS AoD elevation, assuming $R = 3$.

theoretical bound.

Similarly, the modified PREIDG is compared with the ABP method for AoDs elevation θ_r as shown in Figure 3.12. The modified PREIDG shows good performance as compared to the ABP method. Furthermore, the SAGE based high resolution-parameter estimation outperforms and approach close to the theoretical CRLB. The SAGE algorithm marked better performance for the estimation of the complex path gain $(\sqrt{\hat{P_T}}\alpha_r)$ and approaches the CRLB as shown in Figure 3.13. The modified PREIDG and ABP method are unable to estimate the complex-path gains directly. The SAGE algorithm provides estimates of the time-delay τ_r in high-resolution as shown in

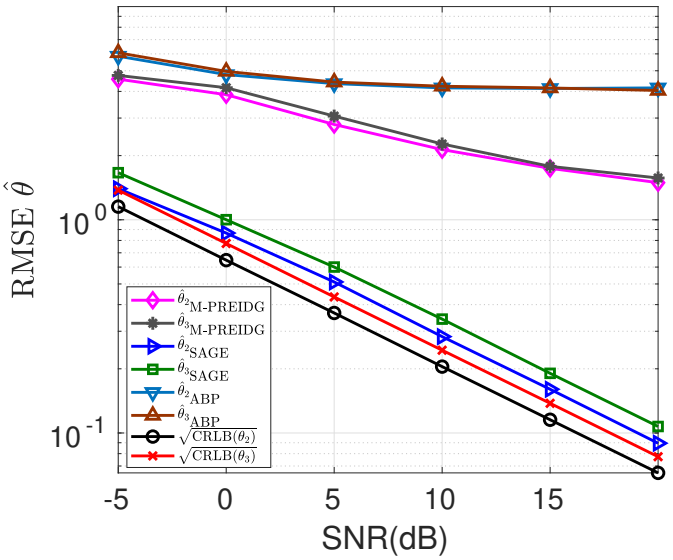


Figure 3.12 – Performance comparison of two-stage and ABP algorithm for NLOS AoDs elevation, assuming $R = 3$.

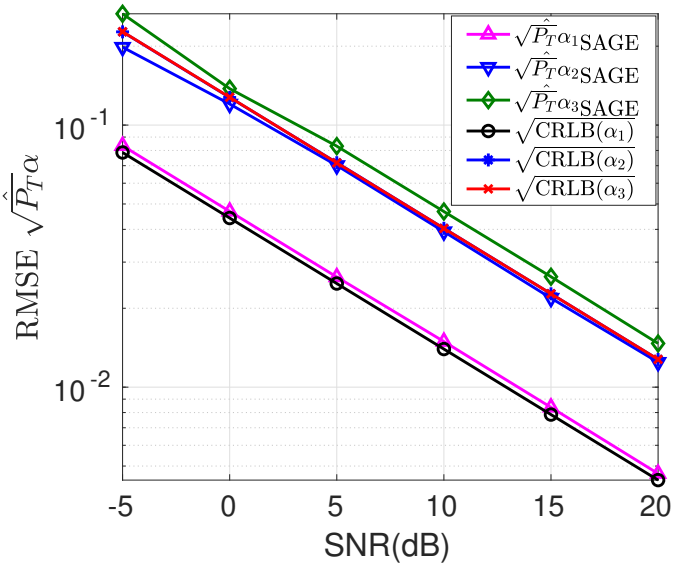


Figure 3.13 – Performance comparison of two-stage and ABP algorithm for $\sqrt{P_T}\alpha_r$.

Figure 3.14. The ABP algorithm is unable to estimate the time-delay τ_r, while the modified PREIDG can only estimate the integer delay. The initialization of the SAGE algorithm based on the coarse estimation achieved by the modified PREIDG enhances the estimation accuracy greatly with reasonable computational complexity.

3.3.10 Exciting URA with 4×4 *and* 8×8 *BMs*

To analyze the 16 available beams formed as the result of Kronecker products of the two 4×4 DFT matrices in terms of finding the azimuth-elevation pair of AoD, i.e., azimuth ϕ_p

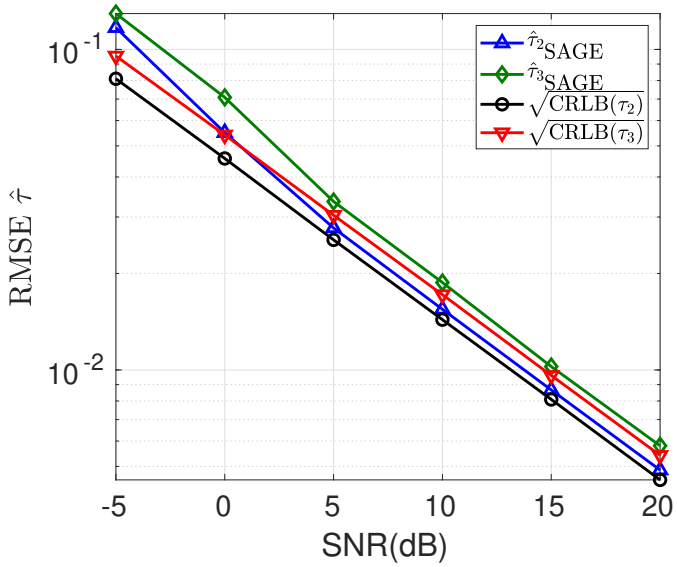


Figure 3.14 – Performance comparison of two-stage algorithm for τ_r

and elevation θ_q, which maximizes the antenna gain, is shown in the Table 3.1. The values are found using equations (3.38) and (3.39). The beams $(3, 7, 11, 15)$ have elevation $\theta_q = 0°$, and therefore the azimuth angles are arbitrary, which can be found numerically using the azimuth range $\phi_p \in \{-90°, 90°\}$ and elevation range $\theta_q \in \{0°, 180°\}$, where $\delta_0 = \nu_0 = 0$, $\delta_1 = \nu_1 = \frac{\pi}{2}$, $\delta_2 = \nu_2 = \pi$, $\delta_3 = \nu_3 = \frac{3\pi}{2}$.

Secondly, for all those beamforming vectors, where the pair of azimuth DFT spatial frequency δ_p and elevation DFT spatial frequency ν_q do not correspond to the pair of azimuth ϕ_p and elevation θ_q angles can be drop out from the channel probing phase, for instance, beams $(10, 12)$. Hence, the beams having arbitrary azimuth ϕ_p and the beams with no azimuth-elevation pairs, i.e., a total of 6 beams will not be used in the channel probing phase. This will help us in reducing pilot overhead.

Next, we are using eight 8×8 BMs to excite an URA having 64 antenna elements with a total number of 64 fixed beams to probe the channel. The details is given in Tables 3.2 and 3.3, respectively.

The beams $(5, 13, 21, 29, 37, 45, 53, 61)$ have elevation $\theta_q = 0°$, and the respective azimuth angles are arbitrary. Secondly, the beams $(28, 30, 34, 35, 36, 38, 39, 40, 44, 46)$ do not have pair of azimuth-elevation spatial frequencies, which will really correspond to the azimuth-elevation pair of angles, and can be found numerically using the azimuth range $\phi_p \in \{-90°, 90°\}$ and elevation range $\theta_q \in \{0°, 180°\}$, where $\delta_0 = \nu_0 = 0$, $\delta_1 = \nu_1 = \frac{\pi}{4}$, $\delta_2 = \nu_2 = \frac{\pi}{2}$, $\delta_3 = \nu_3 = \frac{3\pi}{4}$, $\delta_4 = \nu_4 = \pi$, $\delta_5 = \nu_5 = \frac{5\pi}{4}$, $\delta_6 = \nu_6 = \frac{3\pi}{2}$, $\delta_7 = \nu_7 = \frac{7\pi}{4}$. In summary, there are 10 beams, where there is simply no such azimuth-elevation angle pair for the departing wavefront, which

Table 3.1 – Beamforming angles using eight 4×4 BMs

Serial No	(δ_p, ν_q)	**azimuth** ϕ_p°	**elevation** θ_q°	**Antenna gain**
1	$\mathbf{w}(\delta_0, \nu_0)$	0	90	16
2	$\mathbf{w}(\delta_0, \nu_1)$	0	60	16
3	$\mathbf{w}(\delta_0, \nu_2)$	arbitrary	0	16
4	$\mathbf{w}(\delta_0, \nu_3)$	0	120	16
5	$\mathbf{w}(\delta_1, \nu_0)$	30	90	16
6	$\mathbf{w}(\delta_1, \nu_1)$	35.2644	60	16
7	$\mathbf{w}(\delta_1, \nu_2)$	arbitrary	0	14.6576
8	$\mathbf{w}(\delta_1, \nu_3)$	35.2644	120	16
9	$\mathbf{w}(\delta_2, \nu_0)$	90	90	16
10	$\mathbf{w}(\delta_2, \nu_1)$	no pair, i.e., 90	63	14.6576
11	$\mathbf{w}(\delta_2, \nu_2)$	arbitrary	0	4.7127
12	$\mathbf{w}(\delta_2, \nu_3)$	no pair, i.e., -90	117	14.6576
13	$\mathbf{w}(\delta_3, \nu_0)$	-30	90	16
14	$\mathbf{w}(\delta_3, \nu_1)$	-35.2644	60	16
15	$\mathbf{w}(\delta_3, \nu_2)$	arbitrary	0	14.6576
16	$\mathbf{w}(\delta_3, \nu_3)$	-35.2644	120	16

Table 3.2 – Beamforming angles using eight 8×8 BMs

S.No	(δ_p, ν_q)	**azimuth** ϕ_p°	**elevation** θ_q°	**Antenna gain**
1	$\mathbf{w}(\delta_0, \nu_0)$	0	90	64
2	$\mathbf{w}(\delta_0, \nu_1)$	0	75.5225	64
3	$\mathbf{w}(\delta_0, \nu_2)$	0	60	64
4	$\mathbf{w}(\delta_0, \nu_3)$	0	41.4096	64
5	$\mathbf{w}(\delta_0, \nu_4)$	arbitrary	0	64
6	$\mathbf{w}(\delta_0, \nu_5)$	0	138.5904	64
7	$\mathbf{w}(\delta_0, \nu_6)$	0	120	64
8	$\mathbf{w}(\delta_0, \nu_7)$	0	104.4775	64
9	$\mathbf{w}(\delta_1, \nu_0)$	14.4775	90	64
10	$\mathbf{w}(\delta_1, \nu_1)$	14.9632	75.5225	64
11	$\mathbf{w}(\delta_1, \nu_2)$	16.7787	60	64
12	$\mathbf{w}(\delta_1, \nu_3)$	22.2077	41.4096	64
13	$\mathbf{w}(\delta_1, \nu_4)$	arbitrary	0	62.4411
14	$\mathbf{w}(\delta_1, \nu_5)$	22.2077	138.5904	64
15	$\mathbf{w}(\delta_1, \nu_6)$	16.7787	120	64
16	$\mathbf{w}(\delta_1, \nu_7)$	14.9632	104.4775	64
17	$\mathbf{w}(\delta_2, \nu_0)$	30	90	64
18	$\mathbf{w}(\delta_2, \nu_1)$	31.0909	75.5225	64
19	$\mathbf{w}(\delta_2, \nu_2)$	35.2644	60	64
20	$\mathbf{w}(\delta_2, \nu_3)$	49.1066	41.4096	64

leads to the maximum antenna gain. Therefore, a total number of 18 beams which will not be used in the channel probing phase, and will reduce the pilot overhead.

No-pair example:

Let us take beam-10, i.e., $\mathbf{w}(\delta_2, \nu_1)$ from Table 3.1, Finding the respective azimuth ϕ_p and

elevation θ_q as

$$\hat{\theta}_q = \arccos\left(\frac{\nu_1}{\pi}\right) = 60^\circ, \tag{3.76}$$

where $\nu_1 = \frac{\pi}{2}$. Now using $\hat{\theta}_q$ as

$$\hat{\phi}_p = \arcsin\left(\frac{\delta_2}{\pi \sin\left(\hat{\theta}_q\right)}\right) = \text{no value} \tag{3.77}$$

where $\delta_2 = \pi$.

Table 3.3 – Remaining beamforming angles using eight 8×8 BMs

S.No	(δ_p, ν_q)	**azimuth** ϕ_p°	**elevation** θ_q°	**Antenna gain**
21	$\mathbf{w}(\delta_2, \nu_4)$	arbitrary	0	43.6765
22	$\mathbf{w}(\delta_2, \nu_5)$	49.1066	138.5904	64
23	$\mathbf{w}(\delta_2, \nu_6)$	35.2644	120	64
24	$\mathbf{w}(\delta_2, \nu_7)$	31.0909	104.4775	64
25	$\mathbf{w}(\delta_3, \nu_0)$	48.5904	90	64
26	$\mathbf{w}(\delta_3, \nu_1)$	50.7685	75.5225	64
27	$\mathbf{w}(\delta_3, \nu_2)$	60	60	64
28	$\mathbf{w}(\delta_3, \nu_3)$	no pair,i.e.,90	41.4096(45)	58.1250
29	$\mathbf{w}(\delta_3, \nu_4)$	arbitrary	0	14.6599
30	$\mathbf{w}(\delta_3, \nu_5)$	no pair, i.e., 90	138.5904(135)	58.1250
31	$\mathbf{w}(\delta_3, \nu_6)$	60	120	64
32	$\mathbf{w}(\delta_3, \nu_7)$	50.7685	104.4775	64
33	$\mathbf{w}(\delta_4, \nu_0)$	90	90	64
34	$\mathbf{w}(\delta_4, \nu_1)$	no pair, i.e., -90	75.5225(76)	62.4411
35	$\mathbf{w}(\delta_4, \nu_2)$	no pair,i.e., 90	60(64)	43.6765
36	$\mathbf{w}(\delta_4, \nu_3)$	no pair,i.e., -78	41.4096(41)	14.6559
37	$\mathbf{w}(\delta_4, \nu_4)$	arbitrary	0	8.3486
38	$\mathbf{w}(\delta_4, \nu_5)$	no pair,i.e., -78	138.5904(139)	14.6559
39	$\mathbf{w}(\delta_4, \nu_6)$	no pair,i.e., 90	120(116)	43.6765
40	$\mathbf{w}(\delta_4, \nu_7)$	no pair,i.e., -90	104.4775(104)	62.4411
41	$\mathbf{w}(\delta_5, \nu_0)$	-48.5904	90	64
42	$\mathbf{w}(\delta_5, \nu_1)$	-50.7685	75.5225	64
43	$\mathbf{w}(\delta_5, \nu_2)$	-60	60	64
44	$\mathbf{w}(\delta_5, \nu_3)$	no pair,i.e., -90	41.4096(45)	58.1250
45	$\mathbf{w}(\delta_5, \nu_4)$	arbitrary	0	14.6599
46	$\mathbf{w}(\delta_5, \nu_5)$	no pair,i.e., -90	138.5904(135)	58.1250
47	$\mathbf{w}(\delta_5, \nu_6)$	-60	120	64
48	$\mathbf{w}(\delta_5, \nu_7)$	-50.7685	104.4775	64
49	$\mathbf{w}(\delta_6, \nu_0)$	-30	90	64
50	$\mathbf{w}(\delta_6, \nu_1)$	-31.0909	75.5225	64
51	$\mathbf{w}(\delta_6, \nu_2)$	-35.2644	60	64
52	$\mathbf{w}(\delta_6, \nu_3)$	-49.1066	41.4096	64
53	$\mathbf{w}(\delta_6, \nu_4)$	arbitrary	0	43.6765
54	$\mathbf{w}(\delta_6, \nu_5)$	-49.1066	138.5904	64
55	$\mathbf{w}(\delta_6, \nu_6)$	-35.2644	120	64
56	$\mathbf{w}(\delta_6, \nu_7)$	-31.0909	104.4775	16
57	$\mathbf{w}(\delta_7, \nu_0)$	-14.4775	90	64
58	$\mathbf{w}(\delta_7, \nu_1)$	-14.9632	75.5225	64
59	$\mathbf{w}(\delta_7, \nu_2)$	-16.7787	60	64
60	$\mathbf{w}(\delta_7, \nu_3)$	-22.2077	41.4096	64
61	$\mathbf{w}(\delta_7, \nu_4)$	arbitrary	0	62.441
62	$\mathbf{w}(\delta_7, \nu_5)$	-22.2077	138.5904	64
63	$\mathbf{w}(\delta_7, \nu_6)$	-16.7787	120	64
64	$\mathbf{w}(\delta_7, \nu_7)$	-14.9632	104.4775	16

4 ANALOG AND BASEBAND PRECODING

This chapter deals with the analog-RF and digital baseband precoding assuming the hardware constraints as discussed in chapter 3. In the mobile communication scenario, the receivers/UE's do not have enough degrees of freedom to do a complete job, therefore precoding is introduced in point-to-multipoint connections, to combat the multiuser interference.

This work mainly focuses on a single-cell multiuser scenario where UEs are assumed to operate with a single antenna. We design both analog and baseband precoder for two-dimensional arrays excited by the combination of multiple BMs which produces many fixed Kronecker products of DFT beams. On one hand, it helps in low complexity and energy-efficient implementation but on the other side, it poses a challenge to exploit the channel capacity with a limited number of available analog beams. We propose a strategy for designing an analog precoder while an iteratively WMMSE for sum-utility maximization criteria is derived to design the baseband precoder.

4.1 Overview

mmWave with massive MIMO is a prominent candidate for the next generation of the wireless communication system to rapidly improve the system throughput as less-congested spectrum bands are available. This solution arrives with hardware complexity at the mmWave frequencies which becomes even more challenging in massive MIMO systems [3].
Theoretical studies have shown that for massive MIMO systems, linear precoders achieve near optimal performance [60]. Initial design for analog precoders have focused on low cost PSs [61, 62], keeping in mind, that there exists a performance gap between analog only design and full digital precoding schemes. In the hybrid baseband precoding strategies, to cater for multiuser interference, a baseband precoder in designed in the digital domain. In this work, we use combination of BMs as shown in Figure 3.1, where the PSs are fixed to design the analog RF precoder and design an iterative WMMSE baseband precoder to maximize the overall spectral efficiency.

4.2 Contributions

This chapter is organized as follows

- **System Model:**

We start this chapter with the system model based on the hardware constraints assuming a single cell, single path, and frequency flat-fading channels having multiple single antenna users.

- **Analog precoding:** In this section, assuming the channel state information (CSI) known, we design a novel and simple strategy to design the analog precoder.
- **Baseband precoding:** In this section, we derived an iterative WMMSE solution for our scenario assuming the hardware constraints.
- **Simulation results:**

  Finally, we compare our precoding schemes to the already well-known precoding schemes such as zero forcing (ZF), minimum mean square error (MMSE), and matched filter (MF) precoding techniques to evaluate the performance of our proposed precoding technique.

  The author's research contributions include:

  1. Design of the analog precoder based on the fixed beams generated as the Kronecker products of DFT vectors, to maximize the spectral efficiency of the system.
  2. Derivation of the baseband precoder for our single cell, multiuser scenario using WMMSE algorithm for sum-utility maximization.
  3. Comparison of the propose techniques with the already well known existing techniques to access the performance gap.

4.3 Single-cell, SU-MISO

This section analyzes the case of a single cell, starting by having a single user only, operating with a single antenna.

4.3.1 System Model

The received signal after analog and baseband precoding is given as

$$y = \sqrt{P_T}\mathbf{h}^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}s + n \tag{4.1}$$

Where P_T is the transmit power, $\mathbf{h}^{\mathrm{T}} \in \mathbb{C}^{1\times N}$ is the channel vector between N transmit antennas with N_{RF} chains as $N = M \times N_{RF}$, where M is the number of antennas at each sub-array, Furthermore, all subarrays have the same orientations. The channel $\mathbf{h}$ can be expressed as

$$\mathbf{h} = \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \\ \vdots \\ \mathbf{h}_Z \end{bmatrix} \in \mathbb{C}^{N\times 1} \tag{4.2}$$

where $\mathbf{h}_z$ is the channel for each sub-array and is given as

$$\mathbf{h}_z = \underbrace{\gamma e^{j\varphi_z}}_{\textbf{complex-path gain}} \mathbf{a}(\phi,\theta) \in \mathbb{C}^{M\times 1} \tag{4.3}$$

where $z = 1,\ldots,Z$, $Z = N_{RF}$. γ is the amplitude of LOS path, which is same for each sub-array while the complex phase for LOS path is different for each sub-array and is given as φ_z. $\mathbf{a}(\phi,\theta)$ is the two-dimensional channel steering vector as explained in (3.6). We assume the channel parameters known but it can be estimated as shown in detail in chapter 2 and chapter 3. We consider flat-fading channels for the design of analog and baseband precoder as we assume multi-carrier transmission for data transmission.

$\mathbf{F}_{RF} \in \mathbb{C}^{N\times N_{RF}}$ is the analog precoding matrix which contains columns of the Kronecker product between two DFT matrices. $\mathbf{f}_{BB} \in \mathbb{C}^{N_{RF}\times 1}$ is the digital precoding vector. $s \in \mathbb{C}^{1\times 1}$ is the transmitted symbol.

4.3.2 Analog Precoder Design

In this section, we first design the analog precoder based on the Kronecker product of DFT beams. On one-side, the selection of limited two-dimensional beams improves the ease of implementation in the analog domain while on the other-side compromises the maximum achievable spectral efficiency.

The two-dimensional beamforming vector as defined already in (3.1) and we have a total pool of $M_h M_v = M$, of such beamforming vectors. Now, for a single user, single path scenario, N_{RF} subarrays will be controlled by N_{RF} RF-chains, and will only serve the single user with single path. Therefore, we choose a subset of N_{RF} beamforming vectors from the pool of M beamforming vectors as:

$$\mathbf{w}_1(\delta_p,\nu_q),\ldots,\mathbf{w}_{N_{RF}}(\delta_p,\nu_q) \tag{4.4}$$

where $p = 0,\ldots,M_h-1$ and $q = 0,\ldots,M_v-1$. The LOS path is characterized by its channel vector $\mathbf{h} \in \mathbb{C}^{M\times 1}$ which can be either estimated as given in Chapter 3 or assumed to be known. We

Algorithm 4.1: Analog precoder construction for multi-user case.

Require: $\mathbf{F}_{RF}$ (4.5)
$\mathbf{F}_{RF}$ = Block diagonal empty matrix
for $z = 1$ *to* N_{RF} **do**
  $\mathbf{c}_z = \gamma \mathbf{a}_z(\phi, \theta)$
  $\mathbf{p}_z = \mathbf{c}_z^{\mathrm{H}} [\mathbf{w}_1(\delta_p, \nu_q), \ldots, \mathbf{w}_M(\delta_p, \nu_q)]$
  $j = \max_i \left(\left| [\mathbf{p}_z]_i \right| \right)$
  $\mathbf{w}_z(\delta_p, \nu_q) = \mathbf{w}_j(\delta_p, \nu_q)$
  place the beam $\mathbf{w}_z(\delta_p, \nu_q)$ in the $\mathbf{F}_{RF}$ matrix (4.5)
end
Return $\mathbf{F}_{RF}$

test now the channel vector with every beamforming and select those N_{RF} beamforming vectors, which provides the N_{RF} largest magnitudes of the scalar product. These N_{RF} beamforming vectors are then forming the analog RF precoding matrix. The analog precoding matrix $\mathbf{F}_{RF}$ is structured as

$$\mathbf{F}_{RF} = \begin{bmatrix} \mathbf{w}_1(\delta_p, \nu_q) & \mathbf{0}_M & \ldots & \mathbf{0}_M \\ \mathbf{0}_M & \mathbf{w}_2(\delta_p, \nu_q) & \ldots & \mathbf{0}_M \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_M & \mathbf{0}_M & \ldots & \mathbf{w}_{N_{RF}}(\delta_p, \nu_q) \end{bmatrix} \in \mathbb{C}^{MN_{RF} \times N_{RF}}. \tag{4.5}$$

This is shown in Algorithm 4.1.

4.3.3 Baseband Precoder design

After the selection of near-optimal analog precoding matrix $\mathbf{F}_{RF}$, the baseband precoding vector $\mathbf{f}_{BB}$ for a single user will be designed. The power of the output signal y is written as

$$\begin{aligned} P =& \mathbb{E}\left[yy^{\mathrm{H}}\right] && (4.6)\\ =& \mathbb{E}\left[\left(\sqrt{P_T}\mathbf{h}^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}s + n\right)\left(\sqrt{P_T}\mathbf{h}^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}s + n\right)^{\mathrm{H}}\right] && (4.7)\\ =& \mathbb{E}\Big[P_T\mathbf{h}^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}ss^*\mathbf{f}_{BB}^{\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}^* + \sqrt{P_T}\mathbf{h}^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}sn^* \\ & + \sqrt{P_T}\mathbf{f}_{BB}^{\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}^*s^*n + nn^*\Big] && (4.8)\\ =& P_T\mathbf{h}^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}\mathbb{E}[ss^*]\mathbf{f}_{BB}^{\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}^* + \mathbb{E}\left[nn^*\right] && (4.9)\\ =& P_S + P_N && (4.10) \end{aligned}$$

Where P_S and P_N are signal and noise power. The SNR at the output is

$$\mathrm{SNR} = \frac{P_T \mathbf{h}^{\mathrm{T}} \mathbf{F}_{RF} \mathbf{f}_{BB} \sigma_s^2 \mathbf{f}_{BB}^{\mathrm{H}} \mathbf{F}_{RF}^{\mathrm{H}} \mathbf{h}^*}{\sigma_n^2}. \tag{4.11}$$

The SNR expression (4.11) could be rearranged as Rayleigh quotient and then the expression can be maximized by the eigen vector corresponding the largest eigenvalue of the matrix $\mathbf{Q}$ as

$$\mathrm{SNR} = \mathbf{f}_{BB}^{\mathrm{H}} \mathbf{Q} \mathbf{f}_{BB} \tag{4.12}$$

where $\mathbf{Q}$ is

$$\mathbf{Q} = \frac{P_T \mathbf{F}_{RF}^{\mathrm{H}} \mathbf{h}^* \sigma_s^2 \mathbf{h}^{\mathrm{T}} \mathbf{F}_{RF}}{\sigma_n^2}. \tag{4.13}$$

Therefore $\mathbf{f}_{BB}$ is the eigen vector corresponding to the largest eigenvalue of $\mathbf{Q}$.

4.3.4 Numerical result

In this simulation, we assume a single user having a single antenna served with only one data stream. There are one LOS path, which connect the single user to the BS. The BS has four sub-arrays with four RF-chains $N_{RF} = 4$. We use 10000 channel realizations. The achieved rate is given as

$$R = \log_2 \left(\left| 1 + \frac{P_T \sigma_s^2}{N_s \sigma_n^2} \mathbf{h}^{\mathrm{T}} \mathbf{F}_{RF} \mathbf{f}_{BB} \mathbf{f}_{BB}^{\mathrm{H}} \mathbf{F}_{RF}^{\mathrm{H}} \mathbf{h}^* \right| \right) \tag{4.14}$$

we assume $N_s = 1$, $\sigma_n^2 = \sigma_s^2 = 1$.

Figure 4.1 explains the performance of the proposed method, where the LOS path is chosen randomly as explained in section 3.3.9, i.e., $\phi \sim \mathrm{U}\left(-60^\circ, +60^\circ\right)$ and $\theta \sim \mathrm{U}\left(90^\circ, 130^\circ\right)$, respectively. We construct the analog precoding matrix $\mathbf{F}_{RF}$ as given in Algorithm 4.1. We assume the baseband precoder $\mathbf{f}_{BB}$ as given in section 4.3.3. This is evident that, our proposed method for analog precoding is sub-optimal, as, with SNR of $0\ \mathrm{dB}$, we achieve spectral efficiency close to 2 bits per channel use (bpcu). Moreover for comparison, we construct the analog precoding matrix $\mathbf{F}_{RF}$ with the same beam, i.e., the beam which gives the highest scalar product. The same beam is excited at all subarrays for a single user. It is shown in Figure 4.1 that exciting all subarrays with different beams shows good performance due to the transmission of data with other beams.

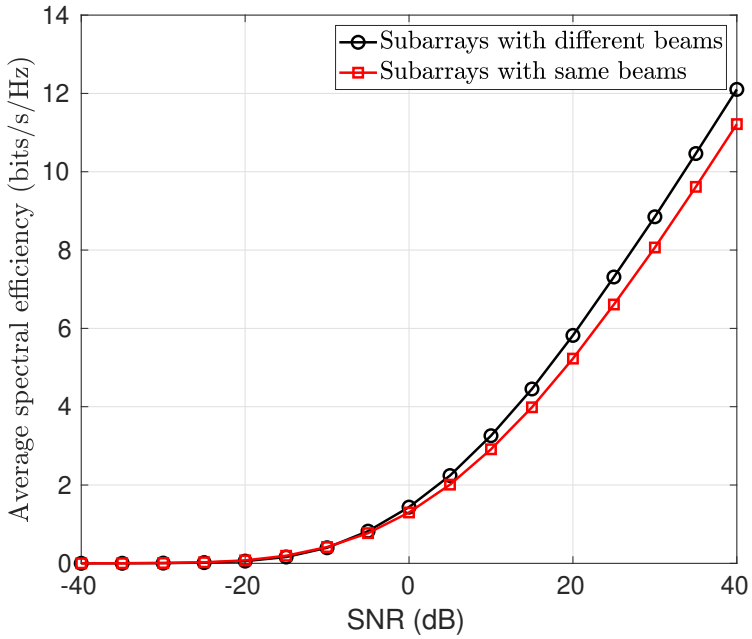


Figure 4.1 – Average spectral efficiency of the proposed hybrid analog and digital precoder.

4.4 Single-cell, MU-MISO

In this section, we design the analog and digital baseband precoder for a multiuser scenario, where each user is equipped with a single antenna. There are K users and each user has a single LOS path. The BS is equipped with two-dimensional subarrays all have the same orientations, excited by combination of BMs termed as PCBMS approach with $N = M \times N_{RF}$ antennas assuming $N_s = N_{RF} = K$, where N_s represents the number of data streams.

4.4.1 System Model

In the downlink scenario, the BS transmits the hybrid beamformed signals to all the users K simultaneously. The received signal vector at all users is expressed as $\mathbf{y} \in \mathbb{C}^{K\times 1}$

$$\mathbf{y} = \sqrt{P_T}\mathbf{H}^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{F}_{BB}\mathbf{s} + \mathbf{n}, \tag{4.15}$$

where $\mathbf{F}_{RF} \in \mathbb{C}^{N\times N_{RF}}$ is the analog precoding matrix, $\mathbf{F}_{BB} \in \mathbb{C}^{N_{RF}\times N_s}$ $\mathbf{s} \in \mathbb{C}^{N_s\times 1}$, $N_s = K = N_{RF}$ is the transmit signal vector satisfying $\mathbb{E}\left[\mathbf{s}\mathbf{s}^{\mathrm{H}}\right] = \sigma_s^2\mathbf{I}_K$ and $\mathbf{n}$ is the complex noise vector with each entry zero mean and σ_n^2 variance as $\mathbb{E}\left[\mathbf{n}\mathbf{n}^{\mathrm{H}}\right] = \sigma_n^2\mathbf{I}_K$. We assume the signals for different users are independent of each other and of the receiver noise. The channel matrix $\mathbf{H} \in \mathbb{C}^{N\times K}$ for all the K users is constructed as

$$\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_K] \in \mathbb{C}^{N\times K} \tag{4.16}$$

where $\mathbf{h}_k, k = 1, \ldots, K$ is the channel of k-th user, given as

$$\mathbf{h}_k = \begin{bmatrix} \mathbf{h}_k^1 \\ \mathbf{h}_k^2 \\ \vdots \\ \mathbf{h}_k^Z \end{bmatrix} \in \mathbb{C}^{N \times 1}, \tag{4.17}$$

where $\mathbf{h}_k^z$ is the channel for each sub-array and is given as

$$\mathbf{h}_k^z = \underbrace{\gamma_k e^{j\varphi_k^z}}_{\textbf{complex-path gain}} \mathbf{a}_k(\phi, \theta) \in \mathbb{C}^{M \times 1} \tag{4.18}$$

where $z = 1, \ldots, Z, Z = N_{RF}$. γ_k is the amplitude for LOS path of the kth user to every subarray. φ_k^z is the phase of the LOS path of the kth user and different for each z-sub-array respectively. $\mathbf{a}_k(\phi, \theta)$ is the two-dimensional channel steering vector of the LOS path of the kth user.

The received signal y_k for the kth user, which contains the target signal, interference from the other $(K-1)$ users and noise is,

$$y_k = \sqrt{P_T}\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^k s_k + \sum_{\substack{j=1 \\ j \neq k}}^{K} \sqrt{P_T}\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^j s_j + n_k \tag{4.19}$$

where $\mathbf{f}_{BB}^k \in \mathbb{C}^{N_{RF} \times 1}$ is the kth column vector of $\mathbf{F}_{BB} \in \mathbb{C}^{N_{RF} \times N_s}$. The received signal-to-interference and noise ratio (SINR) for the kth user can be written as

$$\text{SINR}_k = \frac{P_T \mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^k \sigma_{s,k}^2 \mathbf{f}_{BB}^{k\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^*}{\sum\limits_{\substack{j=1 \\ j \neq k}}^{K} P_T \mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^j \sigma_{s,j}^2 \mathbf{f}_{BB}^{j\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^* + \sigma_n^2} \tag{4.20}$$

The SINR (4.20) expression is derived in Appendix C.1, this can be further re-written as

$$\text{SINR}_k = \frac{\frac{P_T}{K}\left|\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^k\right|^2}{\sum\limits_{\substack{j=1 \\ j \neq k}}^{K} \frac{P_T}{K}\left|\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^j\right|^2 + \sigma_n^2} \tag{4.21}$$

where $\sigma_{s,k} = \frac{1}{K}$. By assuming Gaussian input, the achievable sum-rate of all the K users can be written as

$$R = \sum_{k=1}^{K} \log_2\left(1 + \text{SINR}_k\right). \tag{4.22}$$

4.4.2 Analog Precoder Design

To select the best possible beams in the analog precoding matrix $\mathbf{F}_{RF}$, an exhaustive search method can be used as given in [63]. In our single cell, multi-user scenario, each of the

Algorithm 4.2: Analog precoder construction for multi-user case.

Require: $\mathbf{F}_{RF}$ (4.5)
$\mathbf{F}_{RF}$ = Block diagonal empty matrix
for $k = 1$ *to* N_{RF} **do**
  $\mathbf{c}_k = \gamma_k \mathbf{a}_k(\phi, \theta)$
  $\mathbf{p}_k = \mathbf{c}_k^{\mathrm{H}} [\mathbf{w}_1(\delta_p, \nu_q), \ldots, \mathbf{w}_M(\delta_p, \nu_q)]$
  $j = \max_i \left(\left| [\mathbf{p}_k]_i \right| \right)$
  $\mathbf{w}_k(\delta_p, \nu_q) = \mathbf{w}_j(\delta_p, \nu_q)$
  place the beam $\mathbf{w}_k(\delta_p, \nu_q)$ in the $\mathbf{F}_{RF}$ matrix (4.5)
end
Return $\mathbf{F}_{RF}$

N_{RF} chains will be assigned a beam from the M fixed beams, formed as the Kronecker product of DFT vectors. In the exhaustive search method, $M^{N_{RF}}$ are the total combinations of beams as explained in [63], where we get the optimal combination that maximizes the spectral efficiency R. However, it is highly time inefficient, as it becomes an infeasible solution, either by increasing N_{RF} or increasing M. In our proposed method, the analog precoder is designed by choosing N_{RF} beamforming vectors, one for each user from the set of M fixed beams. Furthermore, each user will be served by one subarray only. We test the channel vector of each user with every beamforming vector and select that beamforming vector, which provides the largest magnitude of the scalar product as shown in Algorithm 4.2

4.4.3 Baseband Precoder Design

To maximize the channel capacity or sum-rate of all the K single-antenna users, we used an iteratively WMMSE approach. We further extend the algorithm to the sum-utility maximization problem as discussed in [64].

Consider a single cell, simultaneously transmit signals to a group of $k = 1, \ldots, K$ single antenna users. The problem of interest is to find the digital baseband precoder $\mathbf{F}_{BB} \in \mathbb{C}^{N_{RF} \times N_s}$ assuming the sub-optimal analog precoder $\mathbf{F}_{RF} \in \mathbb{C}^{N \times N_{RF}}$ fixed, such that a certain utility of the system is maximized by respecting the power budget of the transmitter as given

$$\sum_{k=1}^{K} \mathrm{tr} \left\{ \mathbf{f}_{BB}^{k} \mathbf{f}_{BB}^{k\mathrm{H}} \right\} \leq P_T \tag{4.23}$$

where P_T is the total transmit power budget. Rewriting all the K digital baseband precoding vectors $\mathbf{f}_{BB}^{k}$ in a matrix form as

$$\mathbf{F}_{BB} = \left[\mathbf{f}_{BB}^{1}, \ldots, \mathbf{f}_{BB}^{K} \right] \in \mathbb{C}^{N_{RF} \times N_s} \tag{4.24}$$

4.4.3.1 Weighted Sum-Rate Maximization and a Weighted Sum-MSE Minimization

A popular utility maximization is the weighted sum-rate maximization, which is represented as

$$\mathbf{F}_{BB}^{\text{opt}} = \arg\max_{\mathbf{F}_{BB}} \sum_{k=1}^{K} \beta_k R_k \tag{4.25a}$$

$$\text{subject to} \quad \mathbf{F}_{RF} \in \mathbf{w}_m(\delta_p, \nu_q), m = 0, \ldots, M-1 \tag{4.25b}$$

$$\text{subject to} \quad \text{tr}\left\{\mathbf{F}_{BB}\mathbf{F}_{BB}^{\text{H}}\right\} \leq P_T \tag{4.25c}$$

where the weight β_k is the priority of kth user on the system and R_k is rate of the kth user. The sum-rate for the kth user is given as,

$$R_k = \log_2\left(1 + P_T \mathbf{h}_k^{\text{T}} \mathbf{F}_{RF} \mathbf{f}_{BB}^k \sigma_{s,k}^2 \mathbf{f}_{BB}^{k\text{H}} \mathbf{F}_{RF}^{\text{H}} \mathbf{h}_k^* \left(\sum_{\substack{j=1 \\ j\neq k}}^{K} P_T \mathbf{h}_k^{\text{T}} \mathbf{F}_{RF} \mathbf{f}_{BB}^j \sigma_{s,j}^2 \mathbf{f}_{BB}^{j\text{H}} \mathbf{F}_{RF}^{\text{H}} \mathbf{h}_k^* + \sigma_n^2\right)^{-1}\right). \tag{4.26}$$

Another popular utility maximization problem for MIMO-broadcast channel (BC) is the sum MSE-minimization. Using the independence assumption of s_k and n_k, the MSE for kth user e_k assuming $y_k = \hat{s}_k$ is represented as

$$e_k = \mathbb{E}\left[(\hat{s}_k - s_k)(\hat{s}_k - s_k)^{\text{H}}\right] \tag{4.27}$$

$$\begin{aligned} e_k = &\left(1 - \sqrt{P_T}\mathbf{h}_k^{\text{T}} \mathbf{F}_{RF} \mathbf{f}_{BB}^k \sigma_{s,k}\right)\left(1 - \sqrt{P_T}\mathbf{h}_k^{\text{T}} \mathbf{F}_{RF} \mathbf{f}_{BB}^k \sigma_{s,k}\right)^{\text{H}} \\ &+ \sum_{\substack{j=1 \\ j\neq k}}^{K} P_T \mathbf{h}_k^{\text{T}} \mathbf{F}_{RF} \mathbf{f}_{BB}^j \sigma_{s,j}^2 \mathbf{f}_{BB}^{j\text{H}} \mathbf{F}_{RF}^{\text{H}} \mathbf{h}_k^* + \sigma_{n,k}^2 \end{aligned} \tag{4.28}$$

and therefore the sum-minimization problem can be written as

$$\mathbf{f}_{BB}^{\text{opt}} = \arg\min_{\mathbf{f}_{BB}} \sum_{k=1}^{K} e_k \tag{4.29a}$$

$$\text{subject to} \quad \text{tr}\left\{\mathbf{F}_{BB}\mathbf{F}_{BB}^{\text{H}}\right\} \leq P_T. \tag{4.29b}$$

The expressions for e_k (4.27) are derived in Appendix C.2. Now computing the Lagrangian for this problem as

$$L(\lambda, \mathbf{F}_{BB}) = \sum_{k=1}^{K} e_k + \lambda\left(\sum_{k=1}^{K} \text{tr}\left\{\mathbf{f}_{BB}^k \mathbf{f}_{BB}^{k\text{H}}\right\} - P_T\right) \tag{4.30}$$

where λ is the Lagrangian multiplier for all single antenna users. Let us now use the Karush-Kuhn-Tucker (KKT) conditions.

$$\frac{\partial L\left(\lambda, \mathbf{F}_{BB}\right)}{\partial \lambda} = \sum_{k=1}^{K} \mathrm{tr}\left\{\mathbf{f}_{BB}^{k}\mathbf{f}_{BB}^{k\mathrm{H}}\right\} - P_T = 0. \tag{4.31}$$

To establish equivalence between weighted sum-rate maximization and weighted sum-MSE minimization as mentioned in [64, 65], let $w_k' > 0$ be a scalar weight for kth user, then the problem boils down to

$$\arg \min_{w', \mathbf{f}_{BB}} \sum_{k=1}^{K} \beta_k \left(w_k' e_k - \log w_k'\right) \tag{4.32a}$$

$$\text{subject to} \quad \sum_{k=1}^{K} \mathrm{tr}\left\{\mathbf{f}_{BB}^{k}\mathbf{f}_{BB}^{k\mathrm{H}}\right\} \leq P_T \tag{4.32b}$$

The problem (4.32a), weighted sum-MSE minimization establishes the equivalence between the weighted sum-rate maximization (4.25a), in a sense that the global solution $\mathbf{f}_{BB}^{k}$ for the two problems are identical [64].

Now by fixing the baseband precoder for the kth user $\mathbf{f}_{BB}^{k}$, the objective function (4.32a) is convex with respect to w_k'. Therefore, by checking the first order optimality condition for w_k', we can obtain $w_k'^{\text{opt}}$. Consider the Lagrangian function

$$L\left(\mathbf{w}', \lambda\right) = \sum_{k=1}^{K} \beta_k \left(w_k' e_k - \log w_k'\right) + \lambda \left(\sum_{k=1}^{K} \mathrm{tr}\left\{\mathbf{f}_{BB}^{k}\mathbf{f}_{BB}^{k\mathrm{H}}\right\} - P_T\right) \tag{4.33}$$

where weight vector $\mathbf{w}'$ for all K users can be written as

$$\mathbf{w}' = \left[w_1', \ldots, w_K'\right]^{\mathrm{T}}, \tag{4.34}$$

now computing $\frac{\partial L(\mathbf{w}', \lambda)}{\partial w_k'}$

$$\frac{\partial L\left(\mathbf{w}', \lambda\right)}{\partial w_k'} = \beta_k \left(e_k - w_k'^{-1}\right) = 0 \tag{4.35}$$

$$e_k = w_k'^{-1} \tag{4.36}$$

$$w_k'^{\text{opt}} = e_k^{-1}. \tag{4.37}$$

where e_k is the mean square estimation error and w_k' is a positive weight variable. The equivalence relation simply implies that maximizing sum-rate can be achieved via weighted MSE minimization.

Algorithm 4.3: Baseband precoder design using HBF-WMMSE approach (Pseudo Code).

Require: $\mathbf{F}_{RF}$ (4.5)
Initialize $\mathbf{f}_{BB}^{l}$ randomly
J is the maximum number of iterations for convergence
for $j = 1$ *to* J **do**
    Calculate $w_k^{\prime(j)} = 1/e_k^{(j-1)}$ (4.28)
    Calculate $\mathbf{f}_{BB}^{l(j)}$ (4.38)
end
Return $\mathbf{f}_{BB}^{l\text{opt}}$

4.4.3.2 HBF-WMMSE for sum-utility maximization

In this section, we exploit the equivalence relation discussed in section 4.4.3.1 to derive a simple HBF-WMMSE algorithm for sum-utility maximization problem. Since, the cost function in (4.32a) is convex given the optimization variables $\mathbf{w}', \mathbf{f}_{BB}$. We use block coordinate descent (BCD) method to solve (4.32a) as given in [64]. We therefore, minimize the weighted sum-MSE cost function by sequentially fixing one variable out of two variables, i.e., $\mathbf{w}', \mathbf{F}_{BB}$ and updating the second. To update the w_k' variable, (4.37) is used in the closed form. To update $\mathbf{f}_{BB}^{k}$ we use the following expression

$$\mathbf{f}_{BB}^{l\text{opt}} = \left[\sum_{k=1}^{K} \beta_k w_k' P_T \mathbf{F}_{RF}^{\mathrm{H}} \mathbf{h}_k^{*} \mathbf{h}_k^{\mathrm{T}} \mathbf{F}_{RF} + \lambda \mathbf{I}_K\right]^{-1} \beta_l w_l' \sqrt{P_T} \mathbf{F}_{RF}^{\mathrm{H}} \mathbf{h}_l^{*}, \quad l = 1, \ldots, K. \tag{4.38}$$

Where λ is designed for single cell [65] as

$$\lambda = \frac{\mathrm{tr}\left\{\mathbf{W}'\right\}}{P_T} \tag{4.39}$$

where $\mathbf{W}' = \mathrm{diag}\left\{w_1', \ldots, w_K'\right\}$. The (4.38) is derived in Appendix C.3. The proposed HBF-WMMSE algorithm for the designing of baseband precoder is given in Algorithm 4.3.

4.4.4 Numerical Results

In this section, we present numerical results to evaluate the performance of the proposed HBF-WMMSE algorithm using MU-MISO scenario. The proposed HBF-WMMSE algorithm is compared with classical HBF-MMSE, HBF-ZF and HBF-MF [66, 67]. We assume geometric channel model [47] by employing PCBMS method at BS having URA with $N = N_{RF} \times M = 64$. Each subarray has $M = 16$ antenna elements placed in a squared shape, i.e., 4×4. There are $K = 4$ single antenna users with only LOS paths, and always served by one

subarray, i.e., $K = N_{RF}$. The AoD azimuth ϕ_k and elevation θ_k for LOS path, assuming one sector of a cell as $\phi_k \sim \mathrm{U}(-60°, +60°)$ and $\theta_k \sim \mathrm{U}(90°, 130°)$, respectively. γ_k for LOS is assumed as 1 and φ_k i.e, is the phase of the complex-path gain which is uniformly generated as $\varphi_k \sim \mathrm{U}(0, 2\pi)$. Note, that in PCAPS/PCBMS network, every sub-array has the same magnitude γ_k but different phase φ_k^z. The noise variance is assumed as $\sigma_n^2 = 1$. The weight for the kth user is assumed as $\beta_k = 1$.

Figure 4.2 shows the convergence of the proposed HBF-WMMSE algorithm, since the HBF-WMMSE approach can only converge to local optimum solution, hence, its performance depends on the starting initialization. As our problem is a two-layer optimization, so optimizing analog beamforming matrix $\mathbf{F}_{RF}$ before baseband precoding matrix $\mathbf{F}_{BB}$ can help in faster convergence. The HBF-WMMSE took 3-4 iterations for convergence.

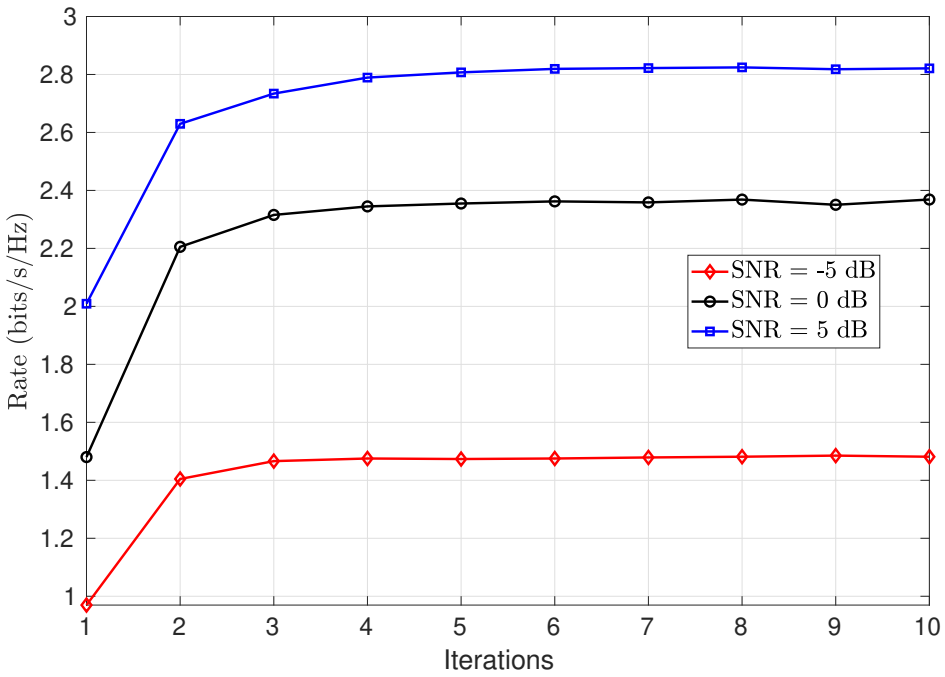


Figure 4.2 – Convergence of the HBF-WMMSE approach.

Next, we analyze the MU-MISO system, where each user has a single antenna. The overall channel matrix of all users will always be a full rank matrix and the rank will be equal to the number of users as $K = N_{RF} = N_s$, where each user is served with one data stream. Hence, the full capacity of the channel is always exploited. As in MISO case, the baseband precoder is designed using HBF-WMMSE without MMSE receiver unlike mentioned in [64, 65], which as a result, ends with a degradation in performance as compared to HBF-MMSE method [66, 67] as shown in Figure 4.3. Note that, in this comparison, the analog precoder is the same for every

method, only the digital baseband precoder is designed differently.

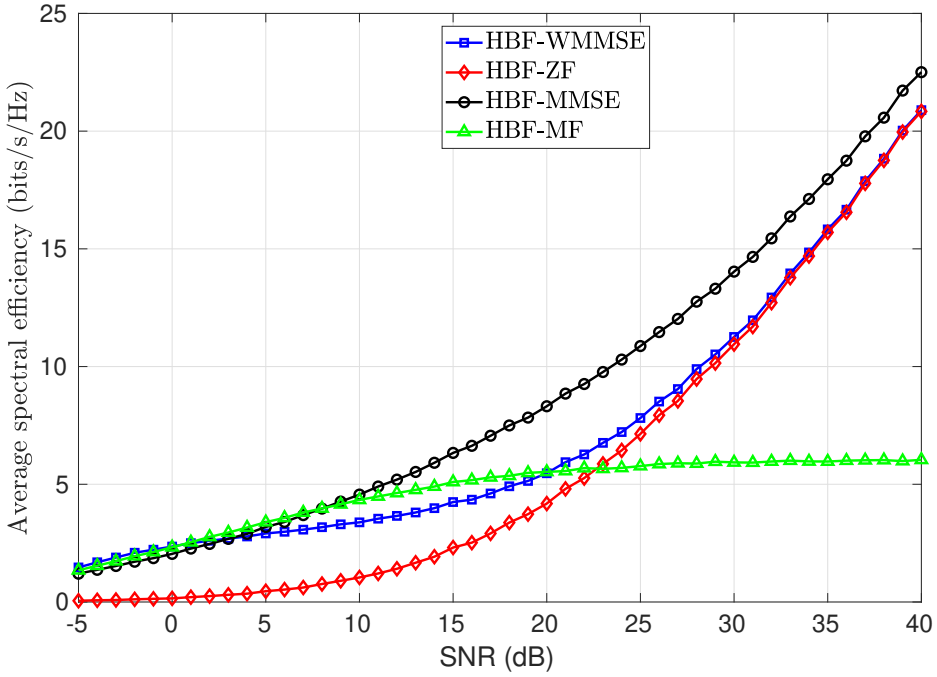


Figure 4.3 – Sum-rate performance comparison for different algorithms as compared to HBF-WMMSE method for MISO system.

Furthermore, in comparison with the FCAPS approach based analog and baseband precoding, it is obvious that FCAPS approach will show improved performance in terms of spectral efficiency due to the formation of narrow beams but degraded performance in terms of energy efficiency.

5 CONCLUSION AND PERSPECTIVES

The next generation of wireless communication systems promises high data rates for the end-users. One way, to make it possible is the introduction of large antenna arrays with mmWaves, which on the side will facilitate the deployment of large antennas but will also provide more bandwidth. The large bandwidth combined with a large number of antenna elements only at the BS poses a challenge of the hardware implementation of such a system, therefore in addition to spectral efficiency, energy efficiency becomes an important design objective. The large scale antenna arrays have multiple engineering challenges including channel estimation. To tackle these important issues, we proposed channel parameter estimation based on BM for ULA and BMs for URA, assuming both frequency flat and selective channels. Finally, based on the this energy efficient setup, a two-step hybrid precoding is proposed.

Specific conclusion of each chapters are given below:

- The channel parameter estimation mentioned in Chapter 2 is based on BM for exciting ULA both for frequency flat and selective channel is discussed. An ML based algorithm is proposed for parameter estimation assuming flat-fading channels. In the second part of the chapter, a two-stage estimation algorithm is proposed to obtain ML estimates of the parameters assuming frequency selective channels. In the first stage, PREIDG algorithm is used for coarse estimation of parameters, which are used to initialize the SAGE algorithm to further refine the estimates. CRLB is derived to assess the performance of ML estimates.
- Chapter 3 is the extension to the two-dimensional channel parameter estimation for frequency selective channels, where an URA is excited using multiple BMs. A two-stage algorithm is proposed to obtain the high-resolution estimates of the parameters. In the first-stage modified PREIDG is proposed to achieve the coarse estimates, which is used to initialize the SAGE algorithm to obtain the high-resolution channel parameter estimates with few iterations.
- Chapter 4 discusses the analog and baseband precoding algorithm for single cell, SU-MISO, and MU-MISO system. On one-hand, the BMs facilitate the energy efficient hardware implementation but on the other-side it pose a challenge on improving the overall spectral efficiency. We proposed a two-stage hybrid precoding algorithm, where in the first stage an analog precoding algorithm is designed, which is then used in the second stage for the designing of the baseband

precoder using HBF-WMMSE.

Future research

As far as future perspectives are concerned, the following are some suggested research lines:

- We can deploy multiple antennas at UE assuming ULA and URA to estimate AoA for frequency selective channels.
- The BS can deploy dual polarized antennas to introduce another degree of freedom.
- The hybrid precoding, we proposed is a two-stage approach, one way which might be investigated is to jointly optimize the analog and baseband precoding. The extension to the multi-cell scenario would be welcome.

REFERENCES

1 PI, Z.; KHAN, F. An introduction to Millimeter-Wave mobile broadband systems. **IEEE Communications Magazine**, v. 49, n. 6, p. 101–107, June 2011. ISSN 0163-6804.

2 RAPPAPORT, T. S. *et al.* Millimeter wave mobile communications for 5G cellular: It will work! **IEEE Access**, v. 1, p. 335–349, 2013. ISSN 2169-3536.

3 RUSEK, F. *et al.* Scaling up MIMO: Opportunities and challenges with very large arrays. **IEEE Signal Processing Magazine**, v. 30, n. 1, p. 40–60, Jan 2013. ISSN 1053-5888.

4 MO, J.; HEATH, R. W. Capacity analysis of one-bit quantized MIMO systems with transmitter channel state information. **IEEE Transactions on Signal Processing**, v. 63, n. 20, p. 5498–5512, Oct 2015. ISSN 1053-587X.

5 Mezghani, A.; Nossek, J. A. On Ultra-Wideband MIMO Systems with 1-bit Quantized Outputs: Performance Analysis and Input Optimization. In: IEEE INTERNATIONAL SYMPOSIUM ON INFORMATION THEORY, 2007, Nice, France. **Proceedings...** Nice, France: IEEE, 2007. p. 1286–1289.

6 Mezghani, A.; Nossek, J. A. Analysis of Rayleigh-fading channels with 1-bit quantized output. In: IEEE INTERNATIONAL SYMPOSIUM ON INFORMATION THEORY, 2008, Toronto, Canada. **Proceedings...** Toronto, Canada: IEEE, 2008. p. 260–264.

7 LEE, C.; CHUNG, W. Hybrid RF-baseband precoding for cooperative multiuser massive MIMO systems with limited RF chains. **IEEE Transactions on Communications**, v. 65, n. 4, p. 1575–1589, April 2017. ISSN 0090-6778.

8 ZHU, J.; XU, W.; WANG, N. Secure massive MIMO systems with limited RF chains. **IEEE Transactions on Vehicular Technology**, v. 66, n. 6, p. 5455–5460, June 2017. ISSN 0018-9545.

9 YANG, L.; ZENG, Y.; ZHANG, R. Efficient channel estimation for Millimeter Wave MIMO with limited RF chains. In: IEEE INTERNATIONAL CONFERENCE ON COMMUNICATIONS, 2016, Kuala Lumpur, Malaysia. **Proceedings...** Kuala Lumpur, Malaysia: IEEE, 2016. p. 1–6.

10 LIAN, L.; LIU, A.; LAU, V. K. N. Optimal-Tuned Weighted LASSO for Massive MIMO Channel Estimation with Limited RF Chains. In: IEEE GLOBAL COMMUNICATIONS CONFERENCE, 2017, Singapore, Singapore. **Proceedings...** Singapore, Singapore: IEEE, 2017. p. 1–6.

11 Moulder, W. F.; Khalil, W.; Volakis, J. L. 60-GHz Two-Dimensionally Scanning Array Employing Wideband Planar Switched Beam Network. **IEEE Antennas and Wireless Propagation Letters**, v. 9, p. 818–821, 2010.

12 Wang, X. *et al.* MIMO Antenna Array System with Integrated 16x16 Butler Matrix and Power Amplifiers for 28GHz Wireless Communication. In: 12TH GERMAN MICROWAVE CONFERENCE (GEMIC), 12., 2019, Stuttgart, Germany. **Proceedings...** Stuttgart, Germany: IEEE, 2019. p. 127–130.

13 Wang, X. *et al.* 28 GHz Multi-Beam Antenna Array based on Wideband High-dimension 16x16 Butler Matrix. In: 13TH EUROPEAN CONFERENCE ON ANTENNAS AND PROPAGATION (EUCAP), 13., 2019, Krakow, Poland. **Proceedings...** Krakow, Poland: IEEE, 2019. p. 1–4.

14 Wang, Y.; Ma, K.; Jian, Z. A low-loss butler matrix using patch element and honeycomb concept on SISL platform. **IEEE Transactions on Microwave Theory and Techniques**, v. 66, n. 8, p. 3622–3631, 2018.

15 Rodriguez-Fernández, J. *et al.* Frequency-Domain Compressive Channel Estimation for Frequency-Selective Hybrid Millimeter Wave MIMO Systems. **IEEE Transactions on Wireless Communications**, v. 17, n. 5, p. 2946–2960, 2018.

16 Wang, B. *et al.* Spatial and Frequency-Wideband Effects in Millimeter-Wave Massive MIMO Systems. **IEEE Transactions on Signal Processing**, v. 66, n. 13, p. 3393–3406, 2018.

17 Wang, B. *et al.* Spatial-Wideband Effect in Massive MIMO with Application in mmWave Systems. **IEEE Communications Magazine**, v. 56, n. 12, p. 134–141, 2018.

18 GURBUZ, A. C.; YAPICI, Y.; GUVENC, I. Sparse Channel Estimation in Millimeter-Wave Communications via Parameter Perturbed OMP. In: IEEE INTERNATIONAL CONFERENCE ON COMMUNICATIONS WORKSHOPS (ICC WORKSHOPS), 13., 2018, Kansas, USA. **Proceedings...** Kansas, USA: IEEE, 2018. p. 1–6.

19 HE, K.; LI, Y.; YIN, C. A novel CS-based non-orthogonal multiple access MIMO system for downlink of MTC in 5G. In: IEEE 28TH ANNUAL INTERNATIONAL SYMPOSIUM ON PERSONAL, INDOOR, AND MOBILE RADIO COMMUNICATIONS (PIMRC), 13., 2017, Montreal, Canada. **Proceedings...** Montreal, Canada: IEEE, 2017. p. 1–5.

20 WANG, P. *et al.* Sparse channel estimation in millimeter wave communications: Exploiting joint AoD-AoA angular spread. In: 2017 IEEE INTERNATIONAL CONFERENCE ON COMMUNICATIONS (ICC), 13., 2017, Paris, France. **Proceedings...** Paris, France: IEEE, 2017. p. 1–6.

21 ELDAR, Y. C.; KUPPINGER, P.; BöLCSKEI, H. Block-sparse signals: Uncertainty relations and efficient recovery. **IEEE Transactions on Signal Processing**, v. 58, n. 6, p. 3042–3054, June 2010. ISSN 1053-587X.

22 ALKHATEEB, A. *et al.* Channel estimation and hybrid precoding for millimeter wave cellular systems. **IEEE Journal of Selected Topics in Signal Processing**, v. 8, n. 5, p. 831–846, Oct 2014. ISSN 1932-4553.

23 LEE, J.; GIL, G.; LEE, Y. H. Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications. **IEEE Transactions on Communications**, v. 64, n. 6, p. 2370–2386, June 2016. ISSN 0090-6778.

24 Venugopal, K. *et al.* Channel estimation for hybrid architecture-based wideband millimeter wave systems. **IEEE Journal on Selected Areas in Communications**, v. 35, n. 9, p. 1996–2009, Sep. 2017. ISSN 1558-0008.

25 SHAHMANSOORI, A. *et al.* Position and orientation estimation through millimeter-wave MIMO in 5G systems. **IEEE Transactions on Wireless Communications**, v. 17, n. 3, p. 1822–1835, March 2018. ISSN 1536-1276.

26 ZHU, D.; CHOI, J.; HEATH, R. W. Auxiliary beam pair design in mmwave cellular systems with hybrid precoding and limited feedback. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING (ICASSP), 41., 2016, Shanghai, China. **Proceedings...** Shanghai, China: IEEE, 2016. p. 7849–7853.

27 ZHU, D.; CHOI, J.; HEATH, R. W. Two-dimensional AoD and AoA acquisition for wideband millimeter-wave systems with dual-polarized MIMO. **IEEE Transactions on Wireless Communications**, v. 16, n. 12, p. 7890–7905, Dec 2017. ISSN 1536-1276.

28 Singh, J.; Ramakrishna, S. On the Feasibility of Codebook-Based Beamforming in Millimeter Wave Systems With Multiple Antenna Arrays. **IEEE Transactions on Wireless Communications**, v. 14, n. 5, p. 2670–2683, 2015.

29 Hu, C. *et al.* Super-Resolution Channel Estimation for MmWave Massive MIMO With Hybrid Precoding. **IEEE Transactions on Vehicular Technology**, v. 67, n. 9, p. 8954–8958, 2018.

30 Ma, W.; Qi, C.; Li, G. Y. High-resolution channel estimation for frequency-selective mmwave massive mimo systems. **IEEE Transactions on Wireless Communications**, v. 19, n. 5, p. 3517–3529, 2020.

31 Sommerkorn, G. *et al.* Reduction of DoA estimation errors caused by antenna array imperfections. In: 29TH EUROPEAN MICROWAVE CONFERENCE, 13., 1999, Munich, Germany. **Proceedings...** Munich, Germany: IEEE, 1999. p. 287–290.

32 EBERHARDT, M.; ESCHLWECH, P.; BIEBL, E. Investigations on antenna array calibration algorithms for direction-of-arrival estimation. **Advances in Radio Science**, v. 14, p. 181–190, 2016. Disponível em: ⟨https://www.adv-radio-sci.net/14/181/2016/⟩.

33 Xuefeng Yin; Fleury, B. H. Nominal direction and direction spread estimation for slightly distributed scatterers using the SAGE algorithm. In: IEEE 61ST VEHICULAR TECHNOLOGY CONFERENCE, 61., 2005, Stockholm, Sweden. **Proceedings...** Stockholm, Sweden: IEEE, 2005. p. 25–29.

34 Feng, R. *et al.* Wireless channel parameter estimation algorithms: Recent advances and future challenges. **China Communications**, v. 15, n. 5, p. 211–228, 2018.

35 de Leon, F. A.; Marciano, J. J. S. Application of MUSIC, ESPRIT and SAGE Algorithms for Narrowband Signal Detection and Localization. In: TENCON 2006 - 2006 IEEE REGION 10 CONFERENCE, 61., 2006, Hong Kong, China. **Proceedings...** Hong Kong, China: IEEE, 2006. p. 25–29.

36 Tschudin, M. *et al.* Comparison between unitary ESPRIT and SAGE for 3-D channel sounding. In: IEEE 49TH VEHICULAR TECHNOLOGY CONFERENCE (CAT. NO.99CH36363), 49., 1999, Houston, USA. **Proceedings...** Houston, USA: IEEE, 1999. p. 1324–1329.

37 Feng, R. *et al.* Comparison of music, unitary ESPRIT, and SAGE algorithms for estimating 3D angles in wireless channels. In: IEEE/CIC INTERNATIONAL CONFERENCE ON COMMUNICATIONS IN CHINA (ICCC), 49., 2017, Qingdao, China. **Proceedings...** Qingdao, China: IEEE, 2017. p. 1–5.

38 Ayach, O. E. *et al.* Spatially Sparse Precoding in Millimeter Wave MIMO Systems. **IEEE Transactions on Wireless Communications**, v. 13, n. 3, p. 1499–1513, 2014.

39 Mendez-Rial, R. *et al.* Hybrid MIMO Architectures for Millimeter Wave Communications: Phase Shifters or Switches? **IEEE Access**, v. 4, p. 247–267, 2016.

40 Zeng, Y.; Zhang, R.; Chen, Z. N. Electromagnetic Lens-Focusing Antenna Enabled Massive MIMO: Performance Improvement and Cost Reduction. **IEEE Journal on Selected Areas in Communications**, v. 32, n. 6, p. 1194–1206, 2014.

41 Lin, T. *et al.* Hybrid Beamforming for Millimeter Wave Systems Using the MMSE Criterion. **IEEE Transactions on Communications**, v. 67, n. 5, p. 3693–3708, 2019.

42 Ardah, K. *et al.* Hybrid Analog-Digital Beamforming Design for SE and EE Maximization in Massive MIMO Networks. **IEEE Transactions on Vehicular Technology**, v. 69, n. 1, p. 377–389, 2020.

43 Ribeiro, L. N. *et al.* Energy Efficiency of mmWave Massive MIMO Precoding With Low-Resolution DACs. **IEEE Journal of Selected Topics in Signal Processing**, v. 12, n. 2, p. 298–312, 2018.

44 Nguyen, D. H. N. *et al.* Hybrid MMSE Precoding and Combining Designs for mmWave Multiuser Systems. **IEEE Access**, v. 5, p. 19167–19181, 2017.

45 Garcia-Rodriguez, A. *et al.* Hybrid analog digital precoding revisited under realistic RF modeling. **IEEE Wireless Communications Letters**, v. 5, n. 5, p. 528–531, 2016.

46 S.M.KAY. **Fundamentals of Statistical Signal processing: Estimation Theory.** [S.l.]: New York, NY, USA: prentice-Hall, 1993, 1993.

47 SAMIMI, M. K.; RAPPAPORT, T. S. 3-D millimeter-wave statistical channel model for 5G wireless system design. **IEEE Transactions on Microwave Theory and Techniques**, v. 64, n. 7, p. 2207–2225, July 2016. ISSN 0018-9480.

48 Jost, T. *et al.* Detection and tracking of mobile propagation channel paths. **IEEE Transactions on Antennas and Propagation**, v. 60, n. 10, p. 4875–4883, 2012.

49 FLEURY, B. *et al.* Channel parameter estimation in mobile radio environments using the SAGE algorithm. **IEEE Journal on Selected Areas in Communications**, v. 17, n. 3, p. 434–450, March 1999. ISSN 0733-8716.

50 Asim, F. *et al.* Maximum likelihood channel estimation for Millimeter-Wave MIMO systems with hybrid beamforming. In: IEEE 23RD INTERNATIONAL ITG WORKSHOP ON SMART ANTENNAS, 2019, Vienna, Austria. **Proceedings...** Vienna, Austria: IEEE, 2019. p. 1–6.

51 Asim, F. *et al.* Channel parameter estimation for millimeter-wave cellular systems with hybrid beamforming. **Signal Processing**, v. 176, p. 2619–2692, July 2020. ISSN 0165-1684.

52 Baggen, L.; Bottcher, M.; Eube, M. 3D-Butler matrix topologies for phased arrays. In: INTERNATIONAL CONFERENCE ON ELECTROMAGNETICS IN ADVANCED APPLICATIONS, 41., 2007, Torino, Italy. **Proceedings...** Torino, Italy: IEEE, 2007. p. 531–534.

53 FRANK, R. L.; ZADOFF, S. A. Phase shift pulse codes with good periodic correlation properties (correspondence). **IRE Trans. on Information Theory**, v. 8, p. 381–382, Oct 1962.

54 POPOVIC, B. M. Generalized chirp-like polyphase sequences with optimum correlation properties. **IEEE Transactions on Information Theory**, v. 38, n. 4, p. 1406–1409, Jul 1992. ISSN 0018-9448.

55 ROHRS, U. H.; LINDE, L. P. Some unique properties and applications of perfect squares minimum phase CAZAC sequences. In: IEEE 1992 SOUTH AFRICAN SYMPOSIUM ON COMMUNICATIONS AND SIGNAL PROCESSING, 1992, Cape Town, South Africa. **Proceedings...** Cape Town, South Africa: IEEE, 1992. p. 155–160.

56 Zoltowski, M. D.; Haardt, M.; Mathews, C. P. Closed-form 2-d angle estimation with rectangular arrays in element space or beamspace via unitary esprit. **IEEE Transactions on Signal Processing**, v. 44, n. 2, p. 316–328, 1996.

57 Zhang, J.; Haardt, M. Channel estimation for hybrid multi-carrier mmwave MIMO systems using three-dimensional unitary esprit in DFT beamspace. In: IEEE 7TH INTERNATIONAL WORKSHOP ON COMPUTATIONAL ADVANCES IN MULTI-SENSOR ADAPTIVE PROCESSING (CAMSAP), 2017, Curacao, Netherlands Antilles. **Proceedings...** Curacao, Netherlands Antilles: IEEE, 2017. p. 1–5.

58 Rakhimov, D. *et al.* Channel Estimation for Hybrid Multi-Carrier mmWave MIMO Systems Using 3-D Unitary Tensor-ESPRIT in DFT beamspace. In: IEEE 53RD ASILOMAR CONFERENCE ON SIGNALS, SYSTEMS, AND COMPUTERS, 53., 2019, Pacific Grove CA, USA. **Proceedings...** Pacific Grove CA, USA: IEEE, 2019. p. 447–451.

59 Fessler, J. A.; Hero, A. O. Space-alternating generalized expectation-maximization algorithm. **IEEE Transactions on Signal Processing**, v. 42, n. 10, p. 2664–2677, 1994.

60 Amadori, P. V.; Masouros, C. Interference-driven antenna selection for massive multiuser mimo. **IEEE Transactions on Vehicular Technology**, v. 65, n. 8, p. 5944–5958, Aug 2016. ISSN 1939-9359.

61 Gholam, F.; Via, J.; Santamaria, I. Beamforming design for simplified analog antenna combining architectures. **IEEE Transactions on Vehicular Technology**, v. 60, n. 5, p. 2373–2378, Jun 2011. ISSN 1939-9359.

62 Venkateswaran, V.; van der Veen, A. Analog beamforming in mimo communications with phase shift networks and online channel estimation. **IEEE Transactions on Signal Processing**, v. 58, n. 8, p. 4131–4143, Aug 2010. ISSN 1941-0476.

63 Han, Y. *et al.* Dft-based hybrid beamforming multiuser systems: Rate analysis and beam selection. **IEEE Journal of Selected Topics in Signal Processing**, v. 12, n. 3, p. 514–528, June 2018. ISSN 1941-0484.

64 Shi, Q. *et al.* An Iteratively Weighted MMSE Approach to Distributed Sum-Utility Maximization for a MIMO Interfering Broadcast Channel. **IEEE Transactions on Signal Processing**, v. 59, n. 9, p. 4331–4340, 2011.

65 Christensen, S. S. *et al.* Weighted sum-rate maximization using weighted mmse for mimo-bc beamforming design. **IEEE Transactions on Wireless Communications**, v. 7, n. 12, p. 4792–4799, 2008.

66 Li, A.; Masouros, C. Hybrid precoding and combining design for millimeter-wave multi-user MIMO based on SVD. In: IEEE INTERNATIONAL CONFERENCE ON COMMUNICATIONS (ICC), 2017, Paris, France. **Proceedings...** Paris, France: IEEE, 2017. p. 1–6.

67 Joham, M. *et al.* Transmit Wiener filter for the downlink of TDDDS-CDMA systems. In: IEEE SEVENTH INTERNATIONAL SYMPOSIUM ON SPREAD SPECTRUM TECHNIQUES

AND APPLICATIONS, 2002, Prague, Czech Republic. **Proceedings...** Prague, Czech Republic: IEEE, 2002. p. 1–5.

68 PAPOULIS, A.; PILLAI, S. U. **Probability, Random Variables, and Stochastic Processes**. [S.l.]: McGraw Hill, 2002.

69 H.V.POOR. **An Introduction to Signal Detection and Estimation** . [S.l.]: Springer Verlagt, 1994.

70 DIETRICH, F. A. **A Tutorial on Channel Estimation with SAGE** . [S.l.]: Technical Report TUM-LNS-TR-06-03, 2006.

71 MCLACHLAN, G. J.; KRISHNAN, T. **The EM Algorithm and Extensions.** . [S.l.]: John Wiley and Sons, Inc., New York, 1997.

APPENDICES

APPENDIX A – ML Estimation and CRLB

We derive the ML estimate for different parameters assuming frequency flat fading channel. We further derive configuration of the SAGE to get the ML estimates of the channel parameters assuming frequency selective channel. We also derive the CRLB to assess the performance of the ML estimates.

A.1 Maximum Likelihood estimation for frequency flat fading channel

Let us assume a random variable $\mathbf{Y}$ which has a multivariate complex Gaussian pdf parameterized by the parameter vector $\boldsymbol{\eta}$ and assuming $L = M$,

$$\mathrm{L}(\mathbf{Y};\boldsymbol{\eta}) = \frac{1}{\pi^{M^2}\det\mathbf{R}}\ \exp\left(-\mathrm{vec}\left\{\mathbf{Y}-\sqrt{P_T}\alpha\mathbf{A}(\mu)\mathbf{C}\right\}^{\mathrm{H}}\mathbf{R}^{-1}\,\mathrm{vec}\left\{\mathbf{Y}-\sqrt{P_T}\alpha\mathbf{A}(\mu)\mathbf{C}\right\}\right) \tag{A.1}$$

where $\boldsymbol{\eta}$

$$\boldsymbol{\eta} = [\sigma_n^2, \sqrt{P_T}\alpha, \mu]^{\mathrm{T}} \tag{A.2}$$

and further assuming spatiality and temporally uncorrelated noise entries ends with the noise variance matrix $\mathbf{R}$

$$\mathbf{R} = \mathrm{E}\left[\mathrm{vec}\{\mathbf{N}\}\ \mathrm{vec}\{\mathbf{N}\}^{\mathrm{H}}\right] = \sigma_n^2\mathbf{I}_{M^2}. \tag{A.3}$$

Now take natural logarithm on both sides of (A.1) as [1]

$$\ell(\mathbf{Y},\boldsymbol{\eta}) = \log_e(\mathrm{L}(\mathbf{Y};\boldsymbol{\eta})) = -M^2\left(\log_e(\pi)+\log_e(\sigma_n^2)\right) - \frac{1}{\sigma_n^2}\mathrm{tr}\left\{(\mathbf{Y}-\sqrt{P_T}\,\alpha\,\mathbf{A}(\mu)\mathbf{C}\right)\left(\mathbf{Y}-\sqrt{P_T}\,\alpha\,\mathbf{A}(\mu)\mathbf{C}\right)^{\mathrm{H}}\Big\} \tag{A.4}$$

The ML estimator can be given as

$$\hat{\boldsymbol{\eta}} = \arg\max_{\boldsymbol{\eta}}\left\{\ell(\mathbf{Y},\boldsymbol{\eta})\right\}. \tag{A.5}$$

Now differentiating (A.4) with respect to σ_n^2 and equating it to zero

$$\frac{\partial\ell(\mathbf{Y},\boldsymbol{\eta})}{\partial\sigma_n^2} = -\frac{M^2}{\sigma_n^2} + \frac{1}{\sigma_n^4}\mathrm{tr}\left\{\left(\mathbf{Y}-\sqrt{P_T}\,\alpha\,\mathbf{A}(\mu)\mathbf{C}\right)\left(\mathbf{Y}-\sqrt{P_T}\,\alpha\,\mathbf{A}(\mu)\mathbf{C}\right)^{\mathrm{H}}\right\} = 0 \tag{A.6}$$

$$\hat{\sigma}_n^2 = \frac{1}{M^2}\mathrm{tr}\left\{\left(\mathbf{Y}-\sqrt{P_T}\,\alpha\,\mathbf{A}(\mu)\mathbf{C}\right)\left(\mathbf{Y}-\sqrt{P_T}\,\alpha\,\mathbf{A}(\mu)\mathbf{C}\right)^{\mathrm{H}}\right\}. \tag{A.7}$$

[1] $\det(c\mathbf{A}) = c^N\det(\mathbf{A})$ if $\mathbf{A}\in\mathbb{C}^{N\times N}$ and $\det(\mathbf{I}_N) = 1$

Now plugging (A.7) into (A.4) leads to

$$\ell\left(\mathbf{Y},\boldsymbol{\eta}\right) = -M^2\log_e(\pi) - M^2\log_e\left(\frac{1}{M^2}\mathrm{tr}\left\{\left(\mathbf{Y}-\sqrt{P_T}\,\alpha\;\mathbf{A}(\mu)\mathbf{C}\right)\left(\mathbf{Y}-\sqrt{P_T}\,\alpha\;\mathbf{A}(\mu)\mathbf{C}\right)^{\mathrm{H}}\right\}\right)$$
$$-\left(\frac{1}{M^2}\mathrm{tr}\left\{\left(\mathbf{Y}-\sqrt{P_T}\,\alpha\;\mathbf{A}(\mu)\mathbf{C}\right)\left(\mathbf{Y}-\sqrt{P_T}\,\alpha\;\mathbf{A}(\mu)\mathbf{C}\right)^{\mathrm{H}}\right\}\right)^{-1}$$
$$\mathrm{tr}\left\{\left(\mathbf{Y}-\sqrt{P_T}\,\alpha\;\mathbf{A}(\mu)\mathbf{C}\right)\left(\mathbf{Y}-\sqrt{P_T}\,\alpha\;\mathbf{A}(\mu)\mathbf{C}\right)^{\mathrm{H}}\right\}. \tag{A.8}$$

Now to maximize the $\ell\left(\mathbf{Y},\boldsymbol{\eta}\right)$, we can first drop any part not depending on $\boldsymbol{\eta}$,

$$\ell\left(\mathbf{Y},\boldsymbol{\eta}\right) = -M^2\log_e\left(\frac{1}{M^2}\mathrm{tr}\left\{\left(\mathbf{Y}-\sqrt{P_T}\,\alpha\;\mathbf{A}(\mu)\mathbf{C}\right)\left(\mathbf{Y}-\sqrt{P_T}\,\alpha\;\mathbf{A}(\mu)\mathbf{C}\right)^{\mathrm{H}}\right\}\right) \tag{A.9}$$

and since $\log_e$ is a monotonic function, we can maximize the so called concentrated loss function

$$\ell_c\left(\mathbf{Y},\boldsymbol{\eta}\right) = -\mathrm{tr}\left\{\left(\mathbf{Y}-\sqrt{P_T}\,\alpha\;\mathbf{A}(\mu)\mathbf{C}\right)\left(\mathbf{Y}-\sqrt{P_T}\,\alpha\;\mathbf{A}(\mu)\mathbf{C}\right)^{\mathrm{H}}\right\} \tag{A.10}$$
$$= -\mathrm{tr}\left\{\mathbf{Y}\mathbf{Y}^{\mathrm{H}}\right\} + \sqrt{P_T}\alpha\mathrm{tr}\left\{\mathbf{A}(\mu)\mathbf{C}\mathbf{Y}^{\mathrm{H}}\right\} + \sqrt{P_T}\alpha^*\mathrm{tr}\left\{\mathbf{Y}\mathbf{C}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}(\mu)\right\}$$
$$-\alpha\alpha^*P_T\mathrm{tr}\left\{\mathbf{A}(\mu)\mathbf{C}\mathbf{C}^{\mathrm{H}}\mathbf{A}(\mu)^{\mathrm{H}}\right\}. \tag{A.11}$$

Now take the derivative of of $\ell_c\left(\mathbf{Y},\boldsymbol{\eta}\right)$ (A.11) with respect to α^*

$$\frac{\partial\ell_c\left(\mathbf{Y},\boldsymbol{\eta}\right)}{\partial\alpha^*} = \sqrt{P_T}\mathrm{tr}\left\{\mathbf{Y}\mathbf{C}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}(\mu)\right\} - \alpha P_T\mathrm{tr}\left\{\mathbf{A}(\mu)\mathbf{C}\mathbf{C}^{\mathrm{H}}\mathbf{A}(\mu)^{\mathrm{H}}\right\} = 0 \tag{A.12}$$

which ends up to

$$(\hat{\sqrt{P_T}\alpha}) = \frac{\mathrm{tr}\left\{\mathbf{Y}\mathbf{C}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}(\mu)\right\}}{\mathrm{tr}\left\{\mathbf{A}(\mu)\mathbf{C}\mathbf{C}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}(\mu)\right\}} = \frac{\mathrm{tr}\left\{\mathbf{Y}\mathbf{C}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}(\mu)\right\}}{M\mathrm{tr}\left\{\mathbf{A}(\mu)\mathbf{A}^{\mathrm{H}}(\mu)\right\}}. \tag{A.13}$$

assuming $\mathbf{C}\mathbf{C}^{\mathrm{H}} = M\mathbf{1}_M$. Now we plug (A.13) into (A.11) and get

$$\ell_c\left(\mathbf{Y},\boldsymbol{\eta}\right) = -\mathrm{tr}\left\{\mathbf{Y}\mathbf{Y}^{\mathrm{H}}\right\} + \frac{\mathrm{tr}\left\{\mathbf{Y}\mathbf{C}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}(\mu)\right\}}{M\mathrm{tr}\left\{\mathbf{A}(\mu)\mathbf{A}^{\mathrm{H}}(\mu)\right\}}\mathrm{tr}\left\{\mathbf{A}(\mu)\mathbf{C}\mathbf{Y}^{\mathrm{H}}\right\}$$
$$+\frac{\mathrm{tr}\left\{\mathbf{A}(\mu)\mathbf{C}\mathbf{Y}^{\mathrm{H}}\right\}}{M\mathrm{tr}\left\{\mathbf{A}(\mu)\mathbf{A}^{\mathrm{H}}(\mu)\right\}}\mathrm{tr}\left\{\mathbf{A}(\mu)\mathbf{C}\mathbf{Y}^{\mathrm{H}}\right\}$$
$$-\frac{\mathrm{tr}\left\{\mathbf{A}(\mu)\mathbf{C}\mathbf{Y}^{\mathrm{H}}\right\}\mathrm{tr}\left\{\mathbf{Y}\mathbf{C}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}(\mu)\right\}}{M^2\left(\mathrm{tr}\left\{\mathbf{A}(\mu)\mathbf{A}^{\mathrm{H}}(\mu)\right\}\right)^2}M\mathrm{tr}\left\{\mathbf{A}(\mu)\mathbf{A}^{\mathrm{H}}(\mu)\right\} \tag{A.14}$$

Now it ends up with

$$\hat{\mu} = \arg\max_{\mu}\ell_c\left(\mathbf{Y},\boldsymbol{\eta}\right) = \arg\max_{\mu}\frac{\mathrm{tr}\left\{\mathbf{A}(\mu)\mathbf{C}\mathbf{Y}^{\mathrm{H}}\right\}\mathrm{tr}\left\{\mathbf{Y}\mathbf{C}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}(\mu)\right\}}{M\mathrm{tr}\left\{\mathbf{A}(\mu)\mathbf{A}^{\mathrm{H}}(\mu)\right\}} \tag{A.15}$$

Now we use $\mathrm{tr}\left\{\mathbf{Y}\mathbf{C}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}(\mu)\right\} = \mathrm{vec}\left\{\mathbf{A}(\mu)\mathbf{C}\right\}^{\mathrm{H}}\mathrm{vec}\left\{\mathbf{Y}\right\}$ to get

$$\hat{\mu} = \arg\max_{\mu}\frac{\mathrm{vec}\left\{\mathbf{A}(\mu)\mathbf{C}\right\}^{\mathrm{H}}\mathrm{vec}\left\{\mathbf{Y}\right\}\mathrm{vec}\left\{\mathbf{Y}\right\}^{\mathrm{H}}\mathrm{vec}\left\{\mathbf{A}(\mu)\mathbf{C}\right\}}{M\mathrm{tr}\left\{\mathbf{A}(\mu)\mathbf{A}^{\mathrm{H}}(\mu)\right\}} \tag{A.16}$$

Now defining $\mathbf{R_Y} = \mathrm{vec}\{\mathbf{Y}\}\,\mathrm{vec}\{\mathbf{Y}\}^{\mathrm{H}}$ as rank-one matrix. The ratio in (A.16) is Rayleigh quotient, which is maximized by the eigenvector of $\mathbf{R_Y}$ corresponding to the largest eigenvalue of $\mathbf{R_Y}$. Since $\mathbf{R_Y}$ is rank-one matrix, there is only one-non-zero eigenvalue and the corresponding eigenvector is $\mathrm{vec}\{\mathbf{Y}\}$. Now let us call this eigenvector $\mathbf{u} = \mathrm{vec}\{\mathbf{Y}\}$ and by reformulating the problem in (A.16) by

$$\begin{aligned}\|\mathbf{u} - \mathrm{vec}\{\mathbf{A}(\mu)\mathbf{C}\}\|_2^2 = \|\mathbf{U} - \mathbf{A}(\mu)\mathbf{C}\|_{\mathrm{F}}^2 = \mathrm{tr}\left\{(\mathbf{U} - \mathbf{A}(\mu)\mathbf{C})\,(\mathbf{U} - \mathbf{A}(\mu)\mathbf{C})^{\mathrm{H}}\right\} \\ = \mathrm{tr}\left\{\mathbf{U}\mathbf{U}^{\mathrm{H}}\right\} - \mathrm{tr}\left\{\mathbf{A}(\mu)\mathbf{C}\mathbf{U}^{\mathrm{H}}\right\} - \mathrm{tr}\left\{\mathbf{U}\mathbf{C}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}(\mu)\right\} + \mathrm{tr}\left\{\mathbf{A}(\mu)\mathbf{C}\mathbf{C}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}(\mu)\right\}. \quad \text{(A.17)}\end{aligned}$$

Now the problem (A.16) can be re-written as

$$\hat{\mu} = \arg\min_{\mu} \|\mathbf{u} - \mathrm{vec}\{\mathbf{A}(\mu)\mathbf{C}\}\|_2^2 \quad \text{(A.18)}$$

therefore we take the derivative of (A.17) with respect to $\mathbf{A}^{\mathrm{H}}(\mu)$ and equating it equal to zero

$$\frac{\partial}{\partial \mathbf{A}^{\mathrm{H}}(\mu)}\left(\mathrm{tr}\left\{\mathbf{U}\mathbf{U}^{\mathrm{H}}\right\} - \mathrm{tr}\left\{\mathbf{A}(\mu)\mathbf{C}\mathbf{U}^{\mathrm{H}}\right\} - \mathrm{tr}\left\{\mathbf{U}\mathbf{C}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}(\mu)\right\} + \mathrm{tr}\left\{\mathbf{A}(\mu)\mathbf{C}\mathbf{C}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}(\mu)\right\}\right) = 0 \quad \text{(A.19)}$$

which leads to

$$-\left(\mathbf{U}\mathbf{C}^{\mathrm{H}}\right)^{\mathrm{T}} + M\mathbf{A}^{\mathrm{T}}(\mu) = 0 \quad \text{(A.20)}$$

$$M\mathbf{A}^{\mathrm{T}}(\mu) = \left(\mathbf{U}\mathbf{C}^{\mathrm{H}}\right)^{\mathrm{T}} \quad \text{(A.21)}$$

since $\mathbf{U} = \mathrm{unvec}\{\mathbf{u}\} = \mathrm{unvec}\{\mathrm{vec}\{\mathbf{Y}\}\} = \mathbf{Y}$ and for notational convenience $\hat{\mathbf{A}}(\mu) = \hat{\mathbf{A}}$, then finally

$$\hat{\mathbf{A}} = \frac{1}{M}\mathbf{Y}\mathbf{C}^{\mathrm{H}} \quad \text{(A.22)}$$

and in order to estimate the spatial frequency $\hat{\mu}$, we can solve this using one dimensional search

$$\hat{\mu} = \arg\min_{\mu} \mathrm{tr}\left\{\mathrm{diag}\left(\hat{\mathbf{A}} - \mathbf{A}(\mu)\right)\mathrm{diag}\left(\hat{\mathbf{A}} - \mathbf{A}(\mu)\right)^{*}\right\}. \quad \text{(A.23)}$$

Using back substituting, $\hat{\mu}$ in (A.13) to get $(\sqrt{\hat{P_T}}\alpha)$. Finally use $\hat{\mu}$ and $(\sqrt{\hat{P_T}}\alpha)$ in (A.7) to estimate σ_n^2.

A.2 Derivation of the FIM $\mathbf{F}(\boldsymbol{\eta})$

The entries of the $\boldsymbol{F}(\boldsymbol{\eta})$ are derive as

$$[\mathbf{F}(\boldsymbol{\eta})]_{11} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{\mathbf{C}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}(\mu)\mathbf{A}(\mu)\mathbf{C}\right\}\right) \tag{A.24}$$

$$[\mathbf{F}(\boldsymbol{\eta})]_{12} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{j\,\mathbf{C}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}(\mu)\mathbf{A}(\mu)\mathbf{C}\right\}\right) \tag{A.25}$$

$$[\mathbf{F}(\boldsymbol{\eta})]_{13} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{\alpha\,\mathbf{C}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}(\mu)\frac{\partial\mathbf{A}(\mu)}{\partial\mu}\mathbf{C}\right\}\right) \tag{A.26}$$

$$[\mathbf{F}(\boldsymbol{\eta})]_{22} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{\mathbf{C}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}(\mu)\mathbf{A}(\mu)\mathbf{C}\right\}\right) \tag{A.27}$$

$$[\mathbf{F}(\boldsymbol{\eta})]_{23} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{-j\,\alpha\,\mathbf{C}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}(\mu)\frac{\partial\mathbf{A}(\mu)}{\partial\mu}\mathbf{C}\right\}\right) \tag{A.28}$$

$$[\mathbf{F}(\boldsymbol{\eta})]_{33} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{|\alpha|^2\mathbf{C}^{\mathrm{H}}\frac{\partial\mathbf{A}^{\mathrm{H}}(\mu)}{\partial\mu}\frac{\partial\mathbf{A}(\mu)}{\partial\mu}\mathbf{C}\right\}\right). \tag{A.29}$$

Derivative of $\mathbf{A}(\mu)$

The partial derivative of $\mathbf{A}(\mu)$ is calculated as

$$\frac{\partial\mathbf{A}(\mu)}{\partial\mu} = \mathrm{diag}\{\mathbf{a}'^{\mathrm{H}}(\mu)\mathbf{w}(\Phi_k)\}_{k=0}^{M-1} \tag{A.30}$$

with

$$\mathbf{a}'(\mu) = \left[0, \quad (-j)\mathrm{e}^{-j\mu}, \ldots, (-j(M-1))\mathrm{e}^{-j(M-1)\mu}\right]^{\mathrm{T}}. \tag{A.31}$$

A.3 SAGE

Let us assume that the observed data $\mathbf{Y}$ as a random variable of the multivariate complex Gaussian pdf denoted by $\mathrm{p}_{\mathbf{Y}}(\mathbf{Y};\boldsymbol{\eta})$ parametrized by the parameter vector $\boldsymbol{\eta}$ can be represented mathematically as

$$\mathrm{p}_{\mathbf{Y}}(\mathbf{Y};\boldsymbol{\eta}) = \frac{1}{\pi^{ML}\det\mathbf{R}}\ \exp\left(-\mathrm{vec}\left\{\mathbf{Y} - \sqrt{P_T}\sum_{r=1}^{R}\alpha_r\mathbf{A}(\mu_r)\mathbf{C}(\tau_r)\right\}^{\mathrm{H}}\right.$$
$$\left.\mathbf{R}^{-1}\ \mathrm{vec}\left\{\mathbf{Y} - \sqrt{P_T}\sum_{r=1}^{R}\alpha_r\mathbf{A}(\mu_r)\mathbf{C}(\tau_r)\right\}\right). \tag{A.32}$$

where $(.)^{\mathrm{H}}$ denotes complex conjugate transposition, $\det$ represents the determinant and $\mathbf{R} \in \mathbb{C}^{ML\times ML}$ is the temporally and spatially uncorrelated noise covariance matrix as

$$\mathbf{R} = \mathrm{E}\left[\mathrm{vec}\{\mathbf{N}\}\ \mathrm{vec}\{\mathbf{N}\}^{\mathrm{H}}\right] = \sigma_n^2\mathbf{I}_{ML} \tag{A.33}$$

where σ_n^2 is the noise variance. The likelihood function $L(\mathbf{Y};\boldsymbol{\eta})$ with respect to parameter vector $\boldsymbol{\eta}$ is given as

$$L(\mathbf{Y};\boldsymbol{\eta}) = \mathrm{p}_{\mathbf{Y}}(\mathbf{Y};\boldsymbol{\eta}) \tag{A.34}$$

The likelihood function $L(\mathbf{Y};\boldsymbol{\eta})$ is a function of the parameter vector, estimation of which is based on a given realization of the random variable, the matrix $\mathbf{Y}$ contains the samples of the complex baseband signals at the antenna array output. The parameter vector $\boldsymbol{\eta}$ can be defined as

$$\boldsymbol{\eta} = \left[\sqrt{P_T}\boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\mu}^{\mathrm{T}}, \boldsymbol{\tau}^{\mathrm{T}}\right]^{\mathrm{T}}. \tag{A.35}$$

The ML estimator can be formed as

$$\hat{\boldsymbol{\eta}} = \arg\max_{\boldsymbol{\eta}} L(\mathbf{Y};\boldsymbol{\eta}) \tag{A.36}$$

(A.36) is the non-linear optimization problem and performing the global maximization directly, consequently does not have a closed-form solution, therefore, the Expectation-Maximization (EM) algorithm gives a sequential approximation of the problem (A.36). EM algorithm performs a sequence of maximization steps in the space of low dimension which ends with the reduction in complexity. But in our scenario, we use the SAGE algorithm for our scenario to solve this non-linear problem with hardware constraints. In the SAGE algorithm, we estimate the channel parameters sequentially unlike EM, where the parameters are estimated in parallel. This could increase the convergence rate as we can use the current estimates from all other multipath, consequently interference subtraction which improves the expectation E-step.

The SAGE algorithm classifies the signals into a complete but an unobservable data $\mathbf{X}$ and an incomplete but an observable data $\mathbf{Y}$. These two complete and incomplete data sets have a relationship which is described by a deterministic many-to-one- mapping as

$$\mathbf{Y} = \boldsymbol{f}(\mathbf{X}) \tag{A.37}$$

Note that, the dimension of parameter vector, i.e., $\dim(\boldsymbol{\eta})$ gives the dimensionality, where $\boldsymbol{\eta}_{\bar{r}}$ contains a all the parameters present in $\boldsymbol{\eta}$ except those contained in $\boldsymbol{\eta}_r$.

A random hidden matrix represented as $\mathbf{X}_r$ having a pdf $\mathrm{p}(\mathbf{X}_r;\boldsymbol{\eta})$ is an admissible hidden data space with respect to $\boldsymbol{\eta}_r$ for $\mathrm{p}(\mathbf{Y};\boldsymbol{\eta})$, if the joint density of $\mathbf{X}_r$ and $\mathbf{Y}$ satisfies the following

$$\mathrm{p}(\mathbf{Y}, \mathbf{X}_r;\boldsymbol{\eta}) = \mathrm{p}(\mathbf{Y} \mid \mathbf{X}_r;\boldsymbol{\eta}_{\bar{r}})\ \mathrm{p}(\mathbf{X}_r;\boldsymbol{\eta}) \tag{A.38}$$

The conditional density of the observation matrix $\mathbf{Y}$ given the hidden matrix $\mathbf{X}_r$ which depends on $\boldsymbol{\eta}_{\bar{r}}$ rather than $\boldsymbol{\eta}_r$ which ultimately means that all information about the channel parameters which is $\boldsymbol{\eta}_r$ is contained in $\mathbf{Y}$ which means in $\mathbf{X}_r$. Furthermore, if $\mathbf{X}_r$ is completely known, then, as a result, $\boldsymbol{\eta}_r$ is not required to be completely known to uniquely define $\mathbf{Y}$. This can be prove using

$$\mathrm{p}(\mathbf{Y};\boldsymbol{\eta}) = \mathrm{p}(\mathbf{Y};\boldsymbol{\eta}_r,\boldsymbol{\eta}_{\bar{r}}). \tag{A.39}$$

Using Bayes-theorem [68] can lead to

$$\mathrm{p}(\mathbf{Y};\boldsymbol{\eta}_r,\boldsymbol{\eta}_{\bar{r}}) = \frac{\mathrm{p}(\mathbf{Y} \mid \mathbf{X}_r;\boldsymbol{\eta}_{\bar{r}})\ \mathrm{p}(\mathbf{X}_r;\boldsymbol{\eta}_r,\boldsymbol{\eta}_{\bar{r}})}{\mathrm{p}(\mathbf{X}_r \mid \mathbf{Y};\boldsymbol{\eta}_r,\boldsymbol{\eta}_{\bar{r}})}, \tag{A.40}$$

now taking the natural logarithm on (A.40) to get the log-likelihood function

$$\ell(\mathbf{Y};\boldsymbol{\eta}_r;\boldsymbol{\eta}_{\bar{r}}) = \ell(\mathbf{Y} \mid \mathbf{X}_r;\boldsymbol{\eta}_{\bar{r}})\ + \ell(\mathbf{X}_r;\boldsymbol{\eta}_r,\boldsymbol{\eta}_{\bar{r}})\ - \ell(\mathbf{X}_r \mid \mathbf{Y};\boldsymbol{\eta}_r,\boldsymbol{\eta}_{\bar{r}}). \tag{A.41}$$

As the hidden data space $\mathbf{X}_r$ is unobservable, the SAGE algorithm solve the problem of maximizing the log-likelihood function $\ell\left(\mathbf{Y};\boldsymbol{\eta}\right)$ by iteratively and sequentially maximizing the conditional expectation of $\ell\left(\mathbf{Y};\boldsymbol{\eta}\right)$ with respect to the hidden data space $\mathbf{X}_r$ given the observed data $\mathbf{Y}$ having the current update of the estimate of parameter vector $\hat{\boldsymbol{\eta}}$. Now take the conditional expectation on both sides of (A.41) with respect to hidden data space $\mathbf{X}_r$ given the $\mathbf{Y}$ and having the current update of the estimate of the parameter vector $\hat{\boldsymbol{\eta}}$ leads to

$$\begin{aligned}\mathrm{E}_{\mathbf{X}_r}\left[\ell(\mathbf{Y};\boldsymbol{\eta}_r,\boldsymbol{\eta}_{\bar{r}} \mid \mathbf{Y};\hat{\boldsymbol{\eta}})\right] &= \mathrm{E}_{\mathbf{X}_r}\left[\ell(\mathbf{Y} \mid \mathbf{X}_r;\boldsymbol{\eta}_{\bar{r}}) \mid \mathbf{Y};\hat{\boldsymbol{\eta}}\right] + \mathrm{E}_{\mathbf{X}_r}\left[\ell(\mathbf{X}_r;\boldsymbol{\eta}_r,\boldsymbol{\eta}_{\bar{r}} \mid \mathbf{Y};\hat{\boldsymbol{\eta}})\right] \\ &- \mathrm{E}_{\mathbf{X}_r}\left[\ell(\mathbf{X}_r \mid \mathbf{Y};\boldsymbol{\eta}_r,\boldsymbol{\eta}_{\bar{r}} \mid \mathbf{Y};\hat{\boldsymbol{\eta}})\right]\end{aligned} \tag{A.42}$$

with

$$\mathrm{E}_{\mathbf{X}_r}\left[\ell(\mathbf{Y};\boldsymbol{\eta}_r,\boldsymbol{\eta}_{\bar{r}} \mid \mathbf{Y};\hat{\boldsymbol{\eta}})\right] = \ell(\mathbf{Y};\boldsymbol{\eta}_r,\boldsymbol{\eta}_{\bar{r}}) \tag{A.43}$$

As we see that the first term on the right hand side of (A.42) is independent on the vector $\boldsymbol{\eta}_r$ and the third term of (A.42) can be derived as

$$\mathrm{E}_{\mathbf{X}_r}\left[\ell(\mathbf{X}_r \mid \mathbf{Y};\boldsymbol{\eta}_r,\boldsymbol{\eta}_{\bar{r}} \mid \mathbf{Y};\hat{\boldsymbol{\eta}})\right] - \mathrm{E}_{\mathbf{X}_r}\left[\ell(\mathbf{X}_r \mid \mathbf{Y};\hat{\boldsymbol{\eta}}) \mid \mathbf{Y};\hat{\boldsymbol{\eta}}\right] \leq 0. \tag{A.44}$$

Now applying Jensen's inequality [69] to (A.44) which states that for a concave function $f(x)$ is given as

$$\mathrm{E}[\boldsymbol{f}(\boldsymbol{x})] \leq \boldsymbol{f}(\mathrm{E}[\boldsymbol{x}]), \tag{A.45}$$

we can prove (A.44) by applying Jensen's inequality as follows

$$\begin{aligned} &= \mathrm{E}_{\mathbf{X}_r}\left[\ln \frac{L(\mathbf{X}_r \mid \mathbf{Y};\boldsymbol{\eta}_r,\boldsymbol{\eta}_{\bar{r}})}{L(\mathbf{X}_r \mid \mathbf{Y};\hat{\boldsymbol{\eta}})} \mid \mathbf{Y};\hat{\boldsymbol{\eta}}\right] \\ &\leq \ln \mathrm{E}_{\mathbf{X}_r}\left[\frac{L(\mathbf{X}_r \mid \mathbf{Y};\boldsymbol{\eta}_r,\boldsymbol{\eta}_{\bar{r}})}{L(\mathbf{X}_r \mid \mathbf{Y};\hat{\boldsymbol{\eta}})} \mid \mathbf{Y};\hat{\boldsymbol{\eta}}\right] \\ &= \ln \int_{X_r} \frac{L(\mathbf{X}_r \mid \mathbf{Y};\boldsymbol{\eta}_r,\boldsymbol{\eta}_{\bar{r}})}{L(\mathbf{X}_r \mid \mathbf{Y};\hat{\boldsymbol{\eta}})} L(\mathbf{X}_r \mid \mathbf{Y};\hat{\boldsymbol{\eta}})\, dX_r \\ &= \ln \int_{X_r} L(\mathbf{X}_r \mid \mathbf{Y};\boldsymbol{\eta}_r,\boldsymbol{\eta}_{\bar{r}})\, dX_r \\ &= \ln(1) = 0. \end{aligned} \tag{A.46}$$

where X_r represents the range of $\mathbf{X}_r$. To maximize (A.42), the only term $\mathrm{E}_{\mathbf{X}_r}\left[\ell(\mathbf{X}_r;\boldsymbol{\eta}_r,\boldsymbol{\eta}_{\bar{r}} \mid \mathbf{Y};\hat{\boldsymbol{\eta}})\right]$ is need to be iteratively and sequentially maximized with respect to $\boldsymbol{\eta}_r$. This can be affirm as

$$\ell(\mathbf{Y};\boldsymbol{\eta}_r,\boldsymbol{\eta}_{\bar{r}}) \geq \ell(\mathbf{Y};\hat{\boldsymbol{\eta}}) \tag{A.47}$$

iff

$$\mathrm{E}_{\mathbf{X}_r}\left[\ell(\mathbf{X}_r;\boldsymbol{\eta}_r,\boldsymbol{\eta}_{\bar{r}} \mid \mathbf{Y};\hat{\boldsymbol{\eta}})\right] \geq \mathrm{E}_{\mathbf{X}_r}\left[\ell(\mathbf{X}_r;\hat{\boldsymbol{\eta}}) \mid \mathbf{Y};\hat{\boldsymbol{\eta}}\right]. \tag{A.48}$$

The complete but unobservable hidden data space is given as

$$\mathbf{X} \hat{=} \left[\mathbf{X}_1,\ldots,\mathbf{X}_R\right] \in \mathbb{C}^{M\times L} \tag{A.49}$$

given as

$$\mathbf{Y} = \boldsymbol{f}(\mathbf{X}) = \boldsymbol{f}\left([\mathbf{X}_1,\ldots,\mathbf{X}_R]\right) = \sum_{r=1}^{R} \mathbf{X}_r \tag{A.50}$$

$$\mathbf{X}_r = \mathbf{S}_r(\boldsymbol{\eta}_r) + \mathbf{N}_r \tag{A.51}$$

$$\mathbf{R}_r = \mathrm{E}[\mathrm{vec}\{\mathbf{N}_r\}\mathrm{vec}\{\mathbf{N}_r\}^{\mathrm{H}}] = \beta_r\sigma_n^2\mathbf{I}_{ML} \tag{A.52}$$

and now the decomposition of the white noise term $\mathbf{N}$

$$\mathbf{N} = \sum_{r=1}^{R} \mathbf{N}_r \tag{A.53}$$

The noise entries of matrix $\mathbf{N}_r$ are temporally and spatially independent having complex Gaussian distributed with variance $\beta_r\sigma_n^2$. The variance of the noise should satisfy the following condition

$$\sum_{r=1}^{R} \beta_r = 1. \tag{A.54}$$

Now let us assume that $\mathbf{X}_r$ are independent with $r = 1, \ldots, R$, we can get

$$\begin{aligned} \ell(\mathbf{Y}; \boldsymbol{\eta}) &= \sum_{r=1}^{R} \ell(\mathbf{X}_r; \boldsymbol{\eta}_r) \\ &= \sum_{r=1}^{R} \ln\left(\frac{1}{(\pi\beta_r\sigma_n^2)^{ML}} \exp\left(\frac{-\|\mathbf{X}_r - \mathbf{S}_r(\boldsymbol{\eta}_r)\|_{\mathrm{F}}^2}{\beta_r\sigma_n^2}\right)\right) \end{aligned} \tag{A.55}$$

and

$$\mathrm{E}_{\mathbf{X}}\left[\ell(\mathbf{X}; \boldsymbol{\eta} \mid \mathbf{Y}; \hat{\boldsymbol{\eta}})\right] = \sum_{r=1}^{R} \mathrm{E}_{\mathbf{X}_r}\left[\ell(\mathbf{X}_r; \boldsymbol{\eta}_r \mid \mathbf{Y}; \hat{\boldsymbol{\eta}})\right]. \tag{A.56}$$

The parameter vector $\boldsymbol{\eta}_r$ which contains the parameters of the channel which parametrize one wavefront/path and each of the R parameter vectors which determines a subset of the main parameter vector $\boldsymbol{\eta}$.

By maximizing $\mathrm{E}_{\mathbf{X}}\left[\ell(\mathbf{X}; \boldsymbol{\eta} \mid \mathbf{Y}; \hat{\boldsymbol{\eta}})\right]$, we can independently maximize each terms of the right hand side of (A.56), as they it depends on different subsets of $\boldsymbol{\eta}$. $\mathrm{E}_{\mathbf{X}_r}\left[\ell(\mathbf{X}_r; \boldsymbol{\eta}_r \mid \mathbf{Y}; \hat{\boldsymbol{\eta}})\right]$ only depends on the parameters of the r-th wavefront/path and therefore can estimate $\hat{\boldsymbol{\eta}}$ of all channel parameters from a previous iteration which simplifies the optimization. The terms in the (A.56) can be written as

$$\mathrm{E}_{\mathbf{X}_r}\left[\ln(\mathbf{X}_r; \boldsymbol{\eta}_r) \mid \mathbf{Y}; \hat{\boldsymbol{\eta}}\right] = \mathrm{E}_{\mathbf{X}_r}\left[\ln b(\mathbf{X}_r) \mid \mathbf{Y}; \hat{\boldsymbol{\eta}}\right] + \boldsymbol{c}^{\mathrm{H}}(\boldsymbol{\eta}_r)\mathrm{E}_{\mathbf{X}_r}\left[\boldsymbol{t}(\mathbf{X}_r) \mid \mathbf{Y}; \hat{\boldsymbol{\eta}}\right] - \ln a(\boldsymbol{\eta}_r) \tag{A.57}$$

To estimate the parameters of one wavefront/path, we identify the pdf of the hidden data space with the exponential family [70, 69, 71] as

$$\mathrm{p}(\mathbf{X}_r; \boldsymbol{\eta}_r) = \frac{b(\mathbf{X}_r)}{a(\boldsymbol{\eta}_r)} \exp(\boldsymbol{c}^{\mathrm{H}}(\boldsymbol{\eta}_r)\boldsymbol{t}(\mathbf{X}_r)) \tag{A.58}$$

The different variables in (A.58) can be identified as

$$a(\boldsymbol{\eta}_r) = \exp\left(\frac{\|\mathrm{vec}\{\mathbf{S}_r(\boldsymbol{\eta}_r)\}\|_2^2}{\beta_r\sigma_n^2}\right) \tag{A.59}$$

$$b(\mathbf{X}_r) = \frac{1}{(\pi\beta_r\sigma_n^2)^{ML}} \exp\left(\frac{-\|\mathrm{vec}\{\mathbf{X}_r\}\|_2^2}{\beta_r\sigma_n^2}\right) \tag{A.60}$$

$$\boldsymbol{c}(\boldsymbol{\eta}_r) = \frac{1}{\beta_r\sigma_n^2} \begin{bmatrix} \mathrm{vec}\{\mathbf{S}_r(\boldsymbol{\eta}_r)\} \\ \mathrm{vec}\{\mathbf{S}_r(\boldsymbol{\eta}_r)^*\} \end{bmatrix} \tag{A.61}$$

$$\boldsymbol{t}(\mathbf{X}_r) = \begin{bmatrix} \mathrm{vec}\{\mathbf{X}_r\} \\ \mathrm{vec}\{\mathbf{X}_r\}^* \end{bmatrix}. \tag{A.62}$$

Keep in mind that $\boldsymbol{t}\left(\mathbf{X}_r\right)$ is a complete sufficient statistic for $\boldsymbol{\eta}_r$. Now the choice of the hidden space introduces wavefront/path by wavefront/path parameter estimation. Now in order to

maximize (A.57) iteratively and sequentially, two steps, i.e., Expectation step E-step and the maximization step M-step are performed. The E-step would estimate the unobservable hidden data space $\hat{\mathbf{X}}_r$. The E-step is actually calculating R expected values of $\boldsymbol{t}(\mathbf{X}_r)$.

E-Step:

The expected value of the second term has to be calculated before maximizing (A.57). It is actually, an estimator for the sufficient statistic for the parameters of the r-th wavefront/path. Therefore, in the E-step the unobservable, complete and hidden data space is estimated conditioned on the observed but incomplete data $\mathbf{Y}$ and also on the previous estimate of the parameter vector $\boldsymbol{\eta}$ which is obtained in the previous iteration as

$$\hat{\mathbf{X}}_r = \mathrm{E}_{\mathbf{X}_r}[\mathbf{X}_r \mid \mathbf{Y}; \hat{\boldsymbol{\eta}}] = \mathbf{S}_r(\hat{\boldsymbol{\eta}}_r) + \beta_r \left(\mathbf{Y} - \sum_{r'=1}^{R} \mathbf{S}_{r'}(\hat{\boldsymbol{\eta}}_{r'}) \right) \tag{A.63}$$

$$\hat{\mathbf{X}}_r = (1-\beta_r)\mathbf{S}_r(\hat{\boldsymbol{\eta}}_r) + \beta_r \left(\mathbf{Y} - \sum_{\substack{r'=1 \\ r' \neq r}}^{R} \mathbf{S}_{r'}(\hat{\boldsymbol{\eta}}_{r'}) \right). \tag{A.64}$$

In (A.64), β_r controls the rate of convergence. When $\beta_r = 0$, the estimate of the hidden data space $\hat{\mathbf{X}}_r$ of the sufficient statistic is not updated for the r-th wavefront/path. Similarly, by choosing $\beta_r = 1$, the convergence rate is largest because of the amount of new information incorporated into the estimate of the hidden data space $\hat{\mathbf{X}}_r$. The stochastic mapping from the hidden data space $\mathbf{X}_r$ to the observable signal $\mathbf{Y}$ is given as

$$\mathbf{Y} = \mathbf{X}_r + \sum_{\substack{r'=1 \\ r' \neq r}}^{R} \mathbf{S}_{r'} + \mathbf{N}_{r'} \tag{A.65}$$

The sequence of the parameter vector $\boldsymbol{\eta}_r(k)$, k represents the number of iterations and is given as

$$\begin{aligned} \boldsymbol{\eta}_r(1) &= [\tau_1] \\ \boldsymbol{\eta}_r(2) &= [\mu_1] \\ \boldsymbol{\eta}_r(3) &= [\alpha_1] \\ \boldsymbol{\eta}_r(4) &= [\tau_2] \\ &\vdots \end{aligned} \tag{A.66}$$

As discussed above, for the fast convergence rate we assume $\beta_r = 1$, which leads the hidden data space estimation as

$$\hat{\mathbf{X}}_r = \mathbf{Y} - \sum_{\substack{r'=1 \\ r' \neq r}}^{R} \mathbf{S}_{r'}(\hat{\boldsymbol{\eta}}_{r'}). \tag{A.67}$$

M-Step:

Using (A.57), (A.59) and (A.61) with the definition of hidden data space of $\hat{\mathbf{X}}_r$ (A.63), we get

$$\hat{\boldsymbol{\eta}}_r = \arg\max_{\eta_r} \left\{ \left(\boldsymbol{c}^{\mathrm{H}}(\boldsymbol{\eta}_r) \mathrm{E}_{\mathbf{X}_r}[\boldsymbol{t}(\mathbf{X}_r) \mid \mathbf{Y}; \hat{\boldsymbol{\eta}}_r] - \ln a(\boldsymbol{\eta}_r) \right) \right\} \tag{A.68}$$

$$\boldsymbol{c}^{\mathrm{H}} = \frac{1}{\beta_r \sigma_n^2} \left[\mathrm{vec}^{\mathrm{H}}\{\mathbf{S}_r(\boldsymbol{\eta}_r)\} \mathrm{vec}^{\mathrm{T}}\{\mathbf{S}_r(\boldsymbol{\eta}_r)\} \right] \tag{A.69}$$

$$\ln a(\boldsymbol{\eta}_r) = \frac{\|\mathrm{vec}\{\mathbf{S}_r(\boldsymbol{\eta}_r)\}\|_2^2}{\beta_r \sigma_n^2} \tag{A.70}$$

$$\mathrm{E}_{\mathbf{X}_r}[\boldsymbol{t}(\mathbf{X}_r) \mid \mathbf{Y}; \hat{\boldsymbol{\eta}}_r] = \left[\mathrm{vec}^{\mathrm{T}}\{\hat{\mathbf{X}}_r\} \mathrm{vec}^{\mathrm{H}}\{\hat{\mathbf{X}}_r\} \right]^{\mathrm{T}} \tag{A.71}$$

$$\hat{\boldsymbol{\eta}}_r = \arg\max_{\eta_r} \left\{ \frac{1}{\beta_r \sigma_n^2} \left(\mathrm{tr}\{\mathbf{S}_r^{\mathrm{H}}(\boldsymbol{\eta}_r) \hat{\mathbf{X}}_r\} + \mathrm{tr}\{\mathbf{S}_r^{\mathrm{T}}(\boldsymbol{\eta}_r) \hat{\mathbf{X}}_r^*\} - \|\mathbf{S}_r(\boldsymbol{\eta}_r)\|_{\mathrm{F}}^2 \right) \right\} \tag{A.72}$$

The first term in right hand side of (A.57) is independent of $\boldsymbol{\eta}_r$ and was omitted. From (2.31) and (A.35), we have [2]

$$\mathbf{S}_r(\boldsymbol{\eta}_r) = \sqrt{P_T} \alpha_r \mathbf{A}(\mu_r) \mathbf{C}(\tau_r), \quad \mathbf{C}(\tau_r) \mathbf{C}^{\mathrm{H}}(\tau_r) = L \mathbf{I}_M \tag{A.73}$$

and with simplification

$$\|\mathbf{S}_r(\boldsymbol{\eta}_r)\|_{\mathrm{F}}^2 = \mathrm{tr}\left\{ \mathbf{S}_r^{\mathrm{H}}(\boldsymbol{\eta}_r) \mathbf{S}_r(\boldsymbol{\eta}_r) \right\} = P_T \alpha_r \alpha_r^* \mathrm{tr}\{\mathbf{C}^{\mathrm{H}}(\tau_r) \mathbf{A}^{\mathrm{H}}(\mu_r) \mathbf{A}(\mu_r) \mathbf{C}(\tau_r)\} \tag{A.74}$$

Putting the values in (A.68) will lead us to

$$\begin{aligned} \hat{\boldsymbol{\eta}}_r &= \arg\max_{\eta_r} \left\{ \frac{1}{\beta_r \sigma_n^2} \left(\sqrt{P_T} \alpha_r^* \mathrm{tr}\{\mathbf{C}^{\mathrm{H}}(\tau_r) \mathbf{A}^{\mathrm{H}}(\mu_r) \hat{\mathbf{X}}_r\} + \sqrt{P_T} \alpha_r \mathrm{tr}\{\mathbf{C}^{\mathrm{T}}(\tau_r) \mathbf{A}^{\mathrm{T}}(\mu_r) \hat{\mathbf{X}}_r^*\} \right. \right. \\ &\quad \left. \left. - P_T \alpha_r \alpha_r^* \mathrm{tr}\{\mathbf{C}^{\mathrm{H}}(\tau_r) \mathbf{A}^{\mathrm{H}}(\mu_r) \mathbf{A}(\mu_r) \mathbf{C}(\tau_r)\} \right) \right\} \qquad \text{(A.75)} \\ &= \arg\max_{\eta_r} \{\boldsymbol{\Lambda}_r(\boldsymbol{\eta}_r)\} \end{aligned}$$

To estimate $\hat{\alpha}_r$, we need to derivate (A.75) and putting it equal to zero,

$$\frac{\partial \boldsymbol{\Lambda}_r(\boldsymbol{\eta}_r)}{\partial \alpha_r^*} = \frac{1}{\beta_r \sigma_n^2} \left(\sqrt{P_T} \mathrm{tr}\left\{ \mathbf{C}^{\mathrm{H}}(\tau_r) \mathbf{A}^{\mathrm{H}}(\mu_r) \hat{\mathbf{X}}_r \right\} - P_T \alpha_r \mathrm{tr}\{\mathbf{C}^{\mathrm{H}}(\tau_r) \mathbf{A}^{\mathrm{H}}(\mu_r) \mathbf{A}(\mu_r) \mathbf{C}(\tau_r)\} \right) = 0 \tag{A.76}$$

[2] $\mathrm{tr}\left(\mathbf{A}^{\mathrm{T}} \mathbf{B}\right) = \mathrm{vec}\left(\mathbf{A}^{\mathrm{T}}\right) \mathrm{vec}\left(\mathbf{B}\right)$

Solving we can get an analytical expression

$$\hat{\alpha}_r = \frac{\mathrm{tr}\left\{\mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r)\hat{\mathbf{X}}_r\right\}}{\sqrt{P_T}\ \mathrm{tr}\{\mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r)\mathbf{A}(\mu_r)\mathbf{C}(\tau_r)\}} \tag{A.77}$$

To obtain an estimator for τ_r and μ_r, put (A.77) in (A.75).

Rewriting (A.75) and solving will lead us to [3]

$$\hat{\boldsymbol{\eta}}_r = \arg\max_{\eta_r}\left\{\frac{1}{\beta_r\sigma_n^2}\left(2\,\frac{\left|\mathrm{tr}\left\{\mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r)\hat{\mathbf{X}}_r\right\}\right|^2}{\mathrm{tr}\{\mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r)\mathbf{A}(\mu_r)\mathbf{C}(\tau_r)\}} - \frac{\left|\mathrm{tr}\left\{\mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r)\hat{\mathbf{X}}_r\right\}\right|^2}{\mathrm{tr}\{\mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r)\mathbf{A}(\mu_r)\mathbf{C}(\tau_r)\}}\right)\right\} \tag{A.78}$$

$$(\hat{\tau}_r, \hat{\mu}_r) = \arg\max_{\tau_r,\mu_r}\left\{\frac{\left|\mathrm{tr}\left\{\mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r)\hat{\mathbf{X}}_r\right\}\right|^2}{\beta_r\sigma_n^2\ \mathrm{tr}\{\mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r)\mathbf{A}(\mu_r)\mathbf{C}(\tau_r)\}}\right\} \tag{A.79}$$

Steps:

The delay estimation $\hat{\tau}_r$ can be iteratively maximized as

$$\hat{\tau}_r = \arg\max_{\tau_r}\left\{\frac{\left|\mathrm{tr}\left\{\mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\hat{\mu}_r)\hat{\mathbf{X}}_r\right\}\right|^2}{\beta_r\sigma_n^2\ \mathrm{tr}\{\mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\hat{\mu}_r)\mathbf{A}(\hat{\mu}_r)\mathbf{C}(\tau_r)\}}\right\}, \tag{A.80}$$

similarly the spatial frequency $\hat{\mu}_r$ can be iteratively maximized as

$$\hat{\mu}_r = \arg\max_{\mu_r}\left\{\frac{\left|\mathrm{tr}\left\{\mathbf{C}^{\mathrm{H}}(\hat{\tau}_r)\mathbf{A}^{\mathrm{H}}(\mu_r)\hat{\mathbf{X}}_r\right\}\right|^2}{\beta_r\sigma_n^2\ \mathrm{tr}\{\mathbf{C}^{\mathrm{H}}(\hat{\tau}_r)\mathbf{A}^{\mathrm{H}}(\mu_r)\mathbf{A}(\mu_r)\mathbf{C}(\hat{\tau}_r)\}}\right\}, \tag{A.81}$$

while in the end $\hat{\sqrt{P_T}}\alpha_r$ can be analytically found as

$$\hat{\sqrt{P_T}}\alpha_r = \frac{\mathrm{tr}\left\{\mathbf{C}^{\mathrm{H}}(\hat{\tau}_r)\mathbf{A}^{\mathrm{H}}(\hat{\mu}_r)\hat{\mathbf{X}}_r\right\}}{\mathrm{tr}\{\mathbf{C}^{\mathrm{H}}(\hat{\tau}_r)\mathbf{A}^{\mathrm{H}}(\hat{\mu}_r)\mathbf{A}(\hat{\mu}_r)\mathbf{C}(\hat{\tau}_r)\}}. \tag{A.82}$$

[3] $2\mathrm{Re}\,(a) = (a + a^*)$

A.4 Derivation of the FIM $\mathbf{F}(\boldsymbol{\eta})$

The entries of the block matrices of the FIM $\mathbf{F}(\boldsymbol{\eta})$ (2.72) are derived below,

$$[\mathbf{F}_{\mathrm{Re}\{\boldsymbol{\alpha}\}\mathrm{Re}\{\boldsymbol{\alpha}\}}]_{ij} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{\sum_{r=1}^{R} \frac{\partial \alpha_r^*}{\partial \mathrm{Re}\{\alpha_i\}} \mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r) \sum_{r=1}^{R} \frac{\partial \alpha_r}{\partial \mathrm{Re}\{\alpha_j\}} \mathbf{A}(\mu_r)\mathbf{C}(\tau_r)\right\}\right) \tag{A.83}$$

$$[\mathbf{F}_{\mathrm{Re}\{\boldsymbol{\alpha}\}\mathrm{Im}\{\boldsymbol{\alpha}\}}]_{ij} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{\sum_{r=1}^{R} \frac{\partial \alpha_r^*}{\partial \mathrm{Re}\{\alpha_i\}} \mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r) \sum_{r=1}^{R} \frac{\partial \alpha_r}{\partial \mathrm{Im}\{\alpha_j\}} \mathbf{A}(\mu_r)\mathbf{C}(\tau_r)\right\}\right) \tag{A.84}$$

$$[\mathbf{F}_{\mathrm{Re}\{\boldsymbol{\alpha}\}\boldsymbol{\mu}}]_{ij} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{\sum_{r=1}^{R} \frac{\partial \alpha_r^*}{\partial \mathrm{Re}\{\alpha_i\}} \mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r) \sum_{r=1}^{R} \alpha_r \frac{\partial \mathbf{A}(\mu_r)}{\partial \mu_j} \mathbf{C}(\tau_r)\right\}\right) \tag{A.85}$$

$$[\mathbf{F}_{\mathrm{Re}\{\boldsymbol{\alpha}\}\boldsymbol{\tau}}]_{ij} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{\sum_{r=1}^{R} \frac{\partial \alpha_r^*}{\partial \mathrm{Re}\{\alpha_i\}} \mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r) \sum_{r=1}^{R} \alpha_r \mathbf{A}(\mu_r) \frac{\partial \mathbf{C}(\tau_r)}{\partial \tau_j}\right\}\right) \tag{A.86}$$

$$[\mathbf{F}_{\mathrm{Im}\{\boldsymbol{\alpha}\}\mathrm{Im}\{\boldsymbol{\alpha}\}}]_{ij} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{\sum_{r=1}^{R} \frac{\partial \alpha_r^*}{\partial \mathrm{Im}\{\alpha_i\}} \mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r) \sum_{r=1}^{R} \frac{\partial \alpha_r}{\partial \mathrm{Im}\{\alpha_j\}} \mathbf{A}(\mu_r)\mathbf{C}(\tau_r)\right\}\right) \tag{A.87}$$

$$[\mathbf{F}_{\mathrm{Im}\{\boldsymbol{\alpha}\}\boldsymbol{\mu}}]_{ij} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{\sum_{r=1}^{R} \frac{\partial \alpha_r^*}{\partial \mathrm{Im}\{\alpha_i\}} \mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r) \sum_{r=1}^{R} \alpha_r \frac{\partial \mathbf{A}(\mu_r)}{\partial \mu_j} \mathbf{C}(\tau_r)\right\}\right) \tag{A.88}$$

$$[\mathbf{F}_{\mathrm{Im}\{\boldsymbol{\alpha}\}\boldsymbol{\tau}}]_{ij} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{\sum_{r=1}^{R} \frac{\partial \alpha_r^*}{\partial \mathrm{Im}\{\alpha_i\}} \mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r) \sum_{r=1}^{R} \alpha_r \mathbf{A}(\mu_r) \frac{\partial \mathbf{C}(\tau_r)}{\partial \tau_j}\right\}\right) \tag{A.89}$$

$$[\mathbf{F}_{\boldsymbol{\mu}\boldsymbol{\mu}}]_{ij} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{\sum_{r=1}^{R} \alpha_r^* \mathbf{C}^{\mathrm{H}}(\tau_r) \frac{\partial \mathbf{A}^{\mathrm{H}}(\mu_r)}{\partial \mu_i} \sum_{r=1}^{R} \alpha_r \frac{\mathbf{A}(\mu_r)}{\partial \mu_j} \mathbf{C}(\tau_r)\right\}\right) \tag{A.90}$$

$$[\mathbf{F}_{\boldsymbol{\mu}\boldsymbol{\tau}}]_{ij} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{\sum_{r=1}^{R} \alpha_r^* \mathbf{C}^{\mathrm{H}}(\tau_r) \frac{\partial \mathbf{A}^{\mathrm{H}}(\mu_r)}{\partial \mu_i} \sum_{r=1}^{R} \alpha_r \mathbf{A}(\mu_r) \frac{\partial \mathbf{C}(\tau_r)}{\partial \tau_j}\right\}\right) \tag{A.91}$$

$$[\mathbf{F}_{\boldsymbol{\tau}\boldsymbol{\tau}}]_{ij} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{\sum_{r=1}^{R} \alpha_r^* \frac{\partial \mathbf{C}^{\mathrm{H}}(\tau_r)}{\partial \tau_i} \mathbf{A}^{\mathrm{H}}(\mu_r) \sum_{r=1}^{R} \alpha_r \mathbf{A}(\mu_r) \frac{\partial \mathbf{C}(\tau_r)}{\partial \tau_j}\right\}\right). \tag{A.92}$$

Derivative of complex-path gain α_r

The partial derivative of α_r and α_r^* is calculated as

$$\frac{\partial \alpha_r}{\partial \mathrm{Re}\{\alpha_i\}} = \frac{\partial \alpha_r^*}{\partial \mathrm{Re}\{\alpha_i\}} = \begin{cases} 1 \text{ if } (r = i) \\ 0 \text{ if } (r \neq i) \end{cases} \tag{A.93}$$

$$\frac{\partial \alpha_r}{\partial \mathrm{Im}\{\alpha_i\}} = -\frac{\partial \alpha_r^*}{\partial \mathrm{Im}\{\alpha_i\}} = \begin{cases} j \text{ if } (r = i) \\ 0 \text{ if } (r \neq i). \end{cases} \tag{A.94}$$

Derivative of $\mathbf{A}(\mu_r)$

The partial derivative of $\mathbf{A}(\mu_r)$ is calculated as

$$\frac{\partial \mathbf{A}(\mu_r)}{\partial \mu_r} = \mathrm{diag}\{\mathbf{a}'^{\mathrm{H}}(\mu_r)\mathbf{w}(\Phi_k)\}_{k=0}^{M-1} \tag{A.95}$$

with

$$\mathbf{a}'(\mu_r) = \left[0, \quad (-j)\mathrm{e}^{-j\mu_r}, \ldots, (-j(M-1))\mathrm{e}^{-j(M-1)\mu_r}\right]^{\mathrm{T}}. \tag{A.96}$$

Derivative of $c(\tau_r)$

The partial derivative of the sequence $c(t)$ with respect to τ_r is calculated as

$$c(t) = \sum_{n=-\infty}^{+\infty} c(n)\, h(t - nT_s) \tag{A.97}$$

where $h(t)$ is the raised cosine (RC) pulse,

$$h(t) = \frac{\sin\left(\pi \frac{t}{T_s}\right)}{\pi \frac{t}{T_s}} \frac{\cos\left(\rho\pi \frac{t}{T_s}\right)}{1 - (2\rho \frac{t}{T_s})^2} \tag{A.98}$$

where $\rho \in [0, 1]$, represents the roll-off factor. The delayed sequence by τ_r is represented as

$$c(t - \tau_r) = \sum_{n=-\infty}^{+\infty} c(n) h(t - nT_s - \tau_r). \tag{A.99}$$

The partial derivative with respect to τ_r can be written as

$$\begin{aligned} \frac{\partial c(t-\tau_r)}{\partial \tau_r} &= \sum_{n=-\infty}^{+\infty} c(n) \frac{\partial h(t - nT_s - \tau_r)}{\partial \tau_r} \\ &= -\sum_{n=-\infty}^{+\infty} \left(c(n)\, \frac{\partial h(\tilde{t})}{\partial \tilde{t}} \,\Big|_{\tilde{t}=t-nT_s-\tau_r} \right). \end{aligned} \tag{A.100}$$

APPENDIX B – High-Resolution Estimation using SAGE and CRLB

We derive and configure the SAGE to achieve high-resolution estimates for the channel parameters. We further derive CRLB to assess the performance of the high-resolution estimates.

B.1 SAGE

We treat the observed data $\mathbf{Y}$ as a random variable with a Gaussian pdf by an unknown parameter vector $\boldsymbol{\eta}$ as

$$\boldsymbol{\eta} = \left[\sqrt{P_T}\boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\psi}^{\mathrm{T}}, \boldsymbol{\mu}^{\mathrm{T}}, \boldsymbol{\tau}^{\mathrm{T}}\right]^{\mathrm{T}} \tag{B.1}$$

where

$$\sqrt{P_T}\boldsymbol{\alpha} = \left[\sqrt{P_T}\alpha_1, \ldots, \sqrt{P_T}\alpha_R\right]^{\mathrm{T}} \in \mathbb{C}^{R\times 1} \tag{B.2}$$

$$\boldsymbol{\psi} = [\psi_1, \ldots, \psi_R]^{\mathrm{T}} \in \mathbb{C}^{R\times 1} \tag{B.3}$$

$$\boldsymbol{\mu} = [\mu_1, \ldots, \mu_R]^{\mathrm{T}} \in \mathbb{C}^{R\times 1} \tag{B.4}$$

$$\boldsymbol{\tau} = [\tau_1, \ldots, \tau_R]^{\mathrm{T}} \in \mathbb{C}^{R\times 1} \tag{B.5}$$

The likelihood can be given as

$$\mathrm{L}(\mathbf{Y};\boldsymbol{\eta}) = \frac{1}{\pi^{ML}\det\mathbf{R}}\ \exp\left(-\mathrm{vec}\left\{\mathbf{Y} - \sqrt{P_T}\sum_{r=1}^{R}\alpha_r\mathbf{A}(\mu_r,\psi_r)\mathbf{C}(\tau_r)\right\}^{\mathrm{H}}\right.$$
$$\left.\mathbf{R}^{-1}\ \mathrm{vec}\left\{\mathbf{Y} - \sqrt{P_T}\sum_{r=1}^{R}\alpha_r\mathbf{A}(\mu_r,\psi_r)\mathbf{C}(\tau_r)\right\}\right). \tag{B.6}$$

SAGE uses the observable but incomplete data space $\mathbf{Y}$ as in (B.6) to find the complete but unobservable data space $\mathbf{X}$ as

$$\mathbf{Y} = \boldsymbol{f}(\mathbf{X}) = \boldsymbol{f}\left([\mathbf{X}_1, \ldots, \mathbf{X}_R]\right) = \sum_{r=1}^{R}\mathbf{X}_r, \tag{B.7}$$

Expectation Step: The conditional expectation of the hidden data space $\mathbf{X}_r$ for each wavefront based on the incomplete but observable data space is given as

$$\hat{\mathbf{X}}_r = \mathbf{E}_{\mathbf{X}_r}[\mathbf{X}_r|\mathbf{Y};\hat{\boldsymbol{\eta}}] = (1-\beta_r)\mathbf{S}_r(\hat{\boldsymbol{\eta}}_r) + \beta_r\left(\mathbf{Y} - \sum_{\substack{r'=1\\ r'\neq r}}^{R}\mathbf{S}_{r'}(\hat{\boldsymbol{\eta}}_r)\right) \tag{B.8}$$

where

$$\mathbf{S}_r(\boldsymbol{\eta}_r) = \sqrt{P_T}\alpha_r \mathbf{A}(\mu_r, \psi_r)\mathbf{C}(\tau_r). \tag{B.9}$$

Assuming $\beta_r = 1$, an estimate of the hidden data space $\hat{\mathbf{X}}_r$ of one wavefront can be given as

$$\hat{\mathbf{X}}_r = \mathbf{Y} - \sum_{\substack{r'=1 \\ r' \neq r}}^{R} \mathbf{S}_{r'}(\hat{\boldsymbol{\eta}}_{r'}). \tag{B.10}$$

Maximization Step:

The cost function can written as

$$\hat{\boldsymbol{\eta}}_r = \arg\max_{\boldsymbol{\eta}_r} \left\{ \frac{1}{\beta_r \sigma_n^2} \left(\mathrm{tr}\{\mathbf{S}_r^{\mathrm{H}}(\boldsymbol{\eta}_r)\hat{\mathbf{X}}_r\} + \mathrm{tr}\{\mathbf{S}_r^{\mathrm{T}}(\boldsymbol{\eta}_r)\hat{\mathbf{X}}_r^*\} - \|\mathbf{S}_r(\boldsymbol{\eta}_r)\|_{\mathrm{F}}^2 \right) \right\} \tag{B.11}$$

where

$$\mathbf{S}_r(\boldsymbol{\eta}_r) = \sqrt{P_T}\alpha_r \mathbf{A}(\mu_r, \psi_r)\mathbf{C}(\tau_r) \tag{B.12}$$

by simplifying

$$\|\mathbf{S_r}(\boldsymbol{\eta}_r)\|_{\mathrm{F}}^2 = \mathrm{tr}\left\{\mathbf{S_r}^{\mathrm{H}}(\boldsymbol{\eta}_r)\mathbf{S_r}(\boldsymbol{\eta_r})\right\} = P_T\alpha_r\alpha_r^* \mathrm{tr}\{\mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r, \psi_r)\mathbf{A}(\mu_r, \psi_r)\mathbf{C}(\tau_r)\}. \tag{B.13}$$

Replacing (B.13) in (B.11), will lead us to

$$\begin{aligned} \hat{\boldsymbol{\eta}}_r &= \arg\max_{\boldsymbol{\eta}_r} \left\{ \frac{1}{\beta_r \sigma_n^2} \left(\sqrt{P_T}\alpha_r^* \mathrm{tr}\{\mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r, \psi_r)\hat{\mathbf{X}}_r\} + \sqrt{P_T}\alpha_r \mathrm{tr}\{\mathbf{C}^{\mathrm{T}}(\tau_r)\mathbf{A}^{\mathrm{T}}(\mu_r, \psi_r)\hat{\mathbf{X}}_r^*\} \right.\right. \\ &\quad \left.\left. - P_T\alpha_r\alpha_r^* \mathrm{tr}\{\mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r, \psi_r)\mathbf{A}(\mu_r, \psi_r)\mathbf{C}(\tau_r)\} \right) \right\} \qquad \text{(B.14)} \\ &= \arg\max_{\eta_r} \{\boldsymbol{\Lambda}_r(\boldsymbol{\eta}_r)\}. \end{aligned}$$

To estimate α_r, we take the derivative of (B.14) with respect to $\sqrt{P_T}\alpha_r^*$ and equate it equal to zero

$$\frac{\partial \boldsymbol{\Lambda}_r(\boldsymbol{\eta}_r)}{\partial \sqrt{P_T}\alpha_r^*} = \frac{1}{\beta_r \sigma_n^2} \left(\mathrm{tr}\left\{ \mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r, \psi_r)\hat{\mathbf{X}}_r \right\} - \sqrt{P_T}\alpha_r \mathrm{tr}\{\mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r, \psi_r)\mathbf{A}(\mu_r, \psi_r)\mathbf{C}(\tau_r)\} \right) = 0. \tag{B.15}$$

Solving (B.15), we arrive with an analytical solution

$$\hat{\sqrt{P_T}}\alpha_r = \frac{\mathrm{tr}\left\{ \mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r, \psi_r)\hat{\mathbf{X}}_r \right\}}{\mathrm{tr}\{\mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r, \psi_r)\mathbf{A}(\mu_r, \psi_r)\mathbf{C}(\tau_r)\}}. \tag{B.16}$$

To further solve for τ_r, μ_r and ψ_r, substituting (B.16) in (B.14) leads to

$$\hat{\boldsymbol{\eta}}_r = \arg\max_{\boldsymbol{\eta}_r} \left\{ \frac{1}{\beta_r \sigma_n^2} \left(\left(\frac{\mathrm{tr}\left\{ \mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r,\psi_r)\hat{\mathbf{X}}_r \right\}}{\mathrm{tr}\{\mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r,\psi_r)\mathbf{A}(\mu_r,\psi_r)\mathbf{C}(\tau_r)\}} \right)^* \mathrm{tr}\left\{ \mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r,\psi_r)\hat{\mathbf{X}}_r \right\} \right.$$
$$+ \frac{\mathrm{tr}\left\{ \mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r,\psi_r)\hat{\mathbf{X}}_r \right\}}{\mathrm{tr}\{\mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r,\psi_r)\mathbf{A}(\mu_r,\psi_r)\mathbf{C}(\tau_r)\}} \mathrm{tr}\left\{ \mathbf{C}^{\mathrm{T}}(\tau_r)\mathbf{A}^{\mathrm{T}}(\mu_r,\psi_r)\hat{\mathbf{X}}_r^* \right\}$$
$$- \left(\frac{\mathrm{tr}\left\{ \mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r,\psi_r)\hat{\mathbf{X}}_r \right\}}{\mathrm{tr}\{\mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r,\psi_r)\mathbf{A}(\mu_r,\psi_r)\mathbf{C}(\tau_r)\}} \right)^* \frac{\mathrm{tr}\left\{ \mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r,\psi_r)\hat{\mathbf{X}}_r \right\}}{\mathrm{tr}\{\mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r,\psi_r)\mathbf{A}(\mu_r,\psi_r)\mathbf{C}(\tau_r)\}}$$
$$\left. \mathrm{tr}\left\{ \mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r,\psi_r)\mathbf{A}(\mu_r,\psi_r)\mathbf{C}(\tau_r) \right\} \right) \right\}. \tag{B.17}$$

We can further write

$$\hat{\boldsymbol{\eta}}_r = \arg\max_{\boldsymbol{\eta}_r} \left\{ \frac{1}{\beta_r \sigma_n^2} \left(\frac{\left| \mathrm{tr}\left\{ \mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r,\psi_r)\hat{\mathbf{X}}_r \right\} \right|^2}{(\mathrm{tr}\{\mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r,\psi_r)\mathbf{A}(\mu_r,\psi_r)\mathbf{C}(\tau_r)\})^*} \right.\right.$$
$$\left.\left. + \frac{\left| \mathrm{tr}\left\{ \mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r,\psi_r)\hat{\mathbf{X}}_r \right\} \right|^2}{\mathrm{tr}\left\{ \mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r,\psi_r)\mathbf{A}(\mu_r,\psi_r)\mathbf{C}(\tau_r) \right\}} - \frac{\left| \mathrm{tr}\left\{ \mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r,\psi_r)\hat{\mathbf{X}}_r \right\} \right|^2}{(\mathrm{tr}\{\mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r,\psi_r)\mathbf{A}(\mu_r,\psi_r)\mathbf{C}(\tau_r)\})^*} \right) \right\}. \tag{B.18}$$

and considering 1-dimensional optimization problems with respect to the sequence of parameter estimates we get

$$\hat{\tau}_r = \arg\max_{\tau_r} \left\{ \frac{\left| \mathrm{tr}\left\{ \mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\hat{\mu}_r,\hat{\psi}_r)\hat{\mathbf{X}}_r \right\} \right|^2}{\beta_r \sigma_n^2 \mathrm{tr}\left\{ \mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\hat{\mu}_r,\hat{\psi}_r)\mathbf{A}(\hat{\mu}_r,\hat{\psi}_r)\mathbf{C}(\tau_r) \right\}} \right\} \tag{B.19}$$

and

$$\hat{\mu}_r = \arg\max_{\mu_r} \left\{ \frac{\left| \mathrm{tr}\left\{ \mathbf{C}^{\mathrm{H}}(\hat{\tau}_r)\mathbf{A}^{\mathrm{H}}(\mu_r,\hat{\psi}_r)\hat{\mathbf{X}}_r \right\} \right|^2}{\beta_r \sigma_n^2 \mathrm{tr}\left\{ \mathbf{C}^{\mathrm{H}}(\hat{\tau}_r)\mathbf{A}^{\mathrm{H}}(\mu_r,\hat{\psi}_r)\mathbf{A}(\mu_r,\hat{\psi}_r)\mathbf{C}(\hat{\tau}_r) \right\}} \right\} \tag{B.20}$$

and

$$\hat{\psi}_r = \arg\max_{\psi_r} \left\{ \frac{\left| \mathrm{tr}\left\{ \mathbf{C}^{\mathrm{H}}(\hat{\tau}_r)\mathbf{A}^{\mathrm{H}}(\hat{\mu}_r,\psi_r)\hat{\mathbf{X}}_r \right\} \right|^2}{\beta_r \sigma_n^2 \mathrm{tr}\left\{ \mathbf{C}^{\mathrm{H}}(\hat{\tau}_r)\mathbf{A}^{\mathrm{H}}(\hat{\mu}_r,\psi_r)\mathbf{A}(\hat{\mu}_r,\psi_r)\mathbf{C}(\hat{\tau}_r) \right\}} \right\}. \tag{B.21}$$

B.2 Derivation of the FIM $\mathbf{F}(\boldsymbol{\eta})$

The entries of the block matrices of the FIM $\mathbf{F}(\boldsymbol{\eta})$ (3.64) are derived below,

$$[\mathbf{F}_{\mathrm{Re}\{\boldsymbol{\alpha}\}\mathrm{Re}\{\boldsymbol{\alpha}\}}]_{ij} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{ \sum_{r=1}^{R} \frac{\partial \alpha_r^*}{\partial \mathrm{Re}\{\alpha_i\}} \mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r,\psi_r) \sum_{r=1}^{R} \frac{\partial \alpha_r}{\partial \mathrm{Re}\{\alpha_j\}} \mathbf{A}(\mu_r,\psi_r)\mathbf{C}(\tau_r) \right\} \right) \tag{B.22}$$

$$[\mathbf{F}_{\mathrm{Re}\{\boldsymbol{\alpha}\}\mathrm{Im}\{\boldsymbol{\alpha}\}}]_{ij} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{ \sum_{r=1}^{R} \frac{\partial \alpha_r^*}{\partial \mathrm{Re}\{\alpha_i\}} \mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r,\psi_r) \sum_{r=1}^{R} \frac{\partial \alpha_r}{\partial \mathrm{Im}\{\alpha_j\}} \mathbf{A}(\mu_r,\psi_r)\mathbf{C}(\tau_r) \right\} \right) \tag{B.23}$$

$$[\mathbf{F}_{\mathrm{Re}\{\boldsymbol{\alpha}\}\boldsymbol{\mu}}]_{ij} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{ \sum_{r=1}^{R} \frac{\partial \alpha_r^*}{\partial \mathrm{Re}\{\alpha_i\}} \mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r,\psi_r) \sum_{r=1}^{R} \alpha_r \frac{\partial \mathbf{A}(\mu_r,\psi_r)}{\partial \mu_j} \mathbf{C}(\tau_r) \right\} \right) \tag{B.24}$$

$$[\mathbf{F}_{\mathrm{Re}\{\boldsymbol{\alpha}\}\boldsymbol{\psi}}]_{ij} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{ \sum_{r=1}^{R} \frac{\partial \alpha_r^*}{\partial \mathrm{Re}\{\alpha_i\}} \mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r,\psi_r) \sum_{r=1}^{R} \alpha_r \frac{\partial \mathbf{A}(\mu_r,\psi_r)}{\partial \psi_j} \mathbf{C}(\tau_r) \right\} \right) \tag{B.25}$$

$$[\mathbf{F}_{\mathrm{Re}\{\boldsymbol{\alpha}\}\boldsymbol{\tau}}]_{ij} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{ \sum_{r=1}^{R} \frac{\partial \alpha_r^*}{\partial \mathrm{Re}\{\alpha_i\}} \mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r,\psi_r) \sum_{r=1}^{R} \alpha_r \mathbf{A}(\mu_r,\psi_r) \frac{\partial \mathbf{C}(\tau_r)}{\partial \tau_j} \right\} \right) \tag{B.26}$$

$$[\mathbf{F}_{\mathrm{Im}\{\boldsymbol{\alpha}\}\mathrm{Im}\{\boldsymbol{\alpha}\}}]_{ij} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{ \sum_{r=1}^{R} \frac{\partial \alpha_r^*}{\partial \mathrm{Im}\{\alpha_i\}} \mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r,\psi_r) \sum_{r=1}^{R} \frac{\partial \alpha_r}{\partial \mathrm{Im}\{\alpha_j\}} \mathbf{A}(\mu_r,\psi_r)\mathbf{C}(\tau_r) \right\} \right) \tag{B.27}$$

$$[\mathbf{F}_{\mathrm{Im}\{\boldsymbol{\alpha}\}\boldsymbol{\mu}}]_{ij} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{ \sum_{r=1}^{R} \frac{\partial \alpha_r^*}{\partial \mathrm{Im}\{\alpha_i\}} \mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r,\psi_r) \sum_{r=1}^{R} \alpha_r \frac{\partial \mathbf{A}(\mu_r,\psi_r)}{\partial \mu_j} \mathbf{C}(\tau_r) \right\} \right) \tag{B.28}$$

$$[\mathbf{F}_{\mathrm{Im}\{\boldsymbol{\alpha}\}\boldsymbol{\psi}}]_{ij} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{ \sum_{r=1}^{R} \frac{\partial \alpha_r^*}{\partial \mathrm{Im}\{\alpha_i\}} \mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r,\psi_r) \sum_{r=1}^{R} \alpha_r \frac{\partial \mathbf{A}(\mu_r,\psi_r)}{\partial \psi_j} \mathbf{C}(\tau_r) \right\} \right) \tag{B.29}$$

$$[\mathbf{F}_{\mathrm{Im}\{\boldsymbol{\alpha}\}\boldsymbol{\tau}}]_{ij} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{ \sum_{r=1}^{R} \frac{\partial \alpha_r^*}{\partial \mathrm{Im}\{\alpha_i\}} \mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r,\psi_r) \sum_{r=1}^{R} \alpha_r \mathbf{A}(\mu_r,\psi_r) \frac{\partial \mathbf{C}(\tau_r)}{\partial \tau_j} \right\} \right) \tag{B.30}$$

$$[\mathbf{F}_{\boldsymbol{\mu\mu}}]_{ij} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{ \sum_{r=1}^{R} \alpha_r^* \mathbf{C}^{\mathrm{H}}(\tau_r) \frac{\partial \mathbf{A}^{\mathrm{H}}(\mu_r,\psi_r)}{\partial \mu_i} \sum_{r=1}^{R} \alpha_r \frac{\mathbf{A}(\mu_r,\psi_r)}{\partial \mu_j} \mathbf{C}(\tau_r) \right\} \right) \tag{B.31}$$

$$[\mathbf{F}_{\boldsymbol{\mu\psi}}]_{ij} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{ \sum_{r=1}^{R} \alpha_r^* \mathbf{C}^{\mathrm{H}}(\tau_r) \frac{\partial \mathbf{A}^{\mathrm{H}}(\mu_r,\psi_r)}{\partial \mu_i} \sum_{r=1}^{R} \alpha_r \frac{\mathbf{A}(\mu_r,\psi_r)}{\partial \psi_j} \mathbf{C}(\tau_r) \right\} \right) \tag{B.32}$$

$$[\mathbf{F}_{\boldsymbol{\mu\tau}}]_{ij} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{ \sum_{r=1}^{R} \alpha_r^* \mathbf{C}^{\mathrm{H}}(\tau_r) \frac{\partial \mathbf{A}^{\mathrm{H}}(\mu_r,\psi_r)}{\partial \mu_i} \sum_{r=1}^{R} \alpha_r \mathbf{A}(\mu_r,\psi_r) \frac{\partial \mathbf{C}(\tau_r)}{\partial \tau_j} \right\} \right) \tag{B.33}$$

$$[\mathbf{F}_{\boldsymbol{\psi\psi}}]_{ij} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{ \sum_{r=1}^{R} \alpha_r^* \mathbf{C}^{\mathrm{H}}(\tau_r) \frac{\partial \mathbf{A}^{\mathrm{H}}(\mu_r,\psi_r)}{\partial \psi_i} \sum_{r=1}^{R} \alpha_r \frac{\mathbf{A}(\mu_r,\psi_r)}{\partial \psi_j} \mathbf{C}(\tau_r) \right\} \right) \tag{B.34}$$

$$[\mathbf{F}_{\boldsymbol{\psi\tau}}]_{ij} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{ \sum_{r=1}^{R} \alpha_r^* \mathbf{C}^{\mathrm{H}}(\tau_r) \frac{\partial \mathbf{A}^{\mathrm{H}}(\mu_r,\psi_r)}{\partial \psi_i} \sum_{r=1}^{R} \alpha_r \mathbf{A}(\mu_r,\psi_r) \frac{\partial \mathbf{C}(\tau_r)}{\partial \tau_j} \right\} \right) \tag{B.35}$$

$$[\mathbf{F}_{\boldsymbol{\tau\tau}}]_{ij} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{ \sum_{r=1}^{R} \alpha_r^* \frac{\partial \mathbf{C}^{\mathrm{H}}(\tau_r)}{\partial \tau_i} \mathbf{A}^{\mathrm{H}}(\mu_r,\psi_r) \sum_{r=1}^{R} \alpha_r \mathbf{A}(\mu_r,\psi_r) \frac{\partial \mathbf{C}(\tau_r)}{\partial \tau_j} \right\} \right). \tag{B.36}$$

Derivative of $\mathbf{A}(\mu_r, \psi_r)$

Given that

$$a_m(\mu_r,\psi_r) = e^{-j[p\mu_r + q\psi_r]}, \tag{B.37}$$

where $m = p + qM_v$, $p \in \{0,\ldots,M_h-1\}$, $q \in \{0,\ldots,M_v-1\}$ and, $m \in \{0,\ldots,M-1\}$ $M = M_h M_v$.

$$\frac{\partial \mathbf{A}(\mu_r,\psi_r)}{\partial \mu_r} = \mathrm{diag}\{\mathbf{a}_h'^{\mathrm{H}}(\mu_r,\psi_r)\mathbf{w}_m(\delta_p,\nu_q)\}_{m=0}^{M-1} \tag{B.38}$$

and

$$\frac{\partial \mathbf{A}(\mu_r,\psi_r)}{\partial \psi_r} = \mathrm{diag}\{\mathbf{a}_v'^{\mathrm{H}}(\mu_r,\psi_r)\mathbf{w}_m(\delta_p,\nu_q)\}_{m=0}^{M-1} \tag{B.39}$$

To get $\mathbf{a}_h'(\mu_r,\psi_r)$ and $\mathbf{a}_v'(\mu_r,\psi_r)$ using (B.37)

$$\mathbf{a}_h'(\mu_r,\psi_r) = \frac{\partial \mathbf{a}(\mu_r,\psi_r)}{\partial \mu_r} = \{-jp\ a_m(\mu_r,\psi_r)\}_{m=0}^{M-1} \tag{B.40}$$

and

$$\mathbf{a}_v'(\mu_r,\psi_r) = \frac{\partial \mathbf{a}(\mu_r,\psi_r)}{\partial \psi_r} = \{-jq\ a_m(\mu_r,\psi_r)\}_{m=0}^{M-1}. \tag{B.41}$$

APPENDIX C – Analog and Baseband Precoders

We derive the achievable sum-rate for single-cell multiuser scenario. We further derive, the digital baseband precoder for WMMSE for sum-utility maximization.

C.1 Achievable sum-rate for single-cell multiuser

After selection of the analog precoding matrix $\mathbf{F}_{RF}$, the power of the output signal y_k for kth user is written as

$$P = \mathbb{E}\left[y_k y_k^{\mathrm{H}}\right] \tag{C.1}$$

now putting terms of y_k (4.19), (C.1) leads to

$$\begin{aligned}
=&\mathbb{E}\left[\left(\sqrt{P_T}\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{k}s_k + \sum_{\substack{j=1\\j\neq k}}^{K}\sqrt{P_T}\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{j}s_j + n_k\right)\right.\\
&\left.\left(\sqrt{P_T}\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{k}s_k + \sum_{\substack{j=1\\j\neq k}}^{K}\sqrt{P_T}\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{j}s_j + n_k\right)^{\mathrm{H}}\right] \qquad \text{(C.2)}\\
=&\mathbb{E}\left[P_T\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{k}s_k s_k^{*}\mathbf{f}_{BB}^{k\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^{*} + \sum_{\substack{j=1\\j\neq k}}^{K}P_T\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{k}s_k s_j^{*}\mathbf{f}_{BB}^{j\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^{*}\right.\\
&+\sqrt{P_T}\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{k}s_k n_k^{*} + \sum_{\substack{j=1\\j\neq k}}^{K}P_T\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{j}s_j s_k^{*}\mathbf{f}_{BB}^{k\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^{*}\\
&+\sum_{\substack{j=1\\j\neq k}}^{K}P_T\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{j}s_j s_j^{*}\mathbf{f}_{BB}^{j\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^{*} + \sum_{\substack{j=1\\j\neq k}}^{K}\sqrt{P_T}\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{j}s_j n_k^{*}\\
&\left.+\sqrt{P_T}\mathbf{f}_{BB}^{k\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^{*}s_k^{*}n_k + \sum_{\substack{j=1\\j\neq k}}^{K}\sqrt{P_T}\mathbf{f}_{BB}^{j\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^{*}s_j^{*}n_k + n_k n_k^{*}\right] \qquad \text{(C.3)}
\end{aligned}$$

Now by applying expectation inside and solving,

$$
\begin{aligned}
&= P_T \mathbf{h}_k^{\mathrm{T}} \mathbf{F}_{RF} \mathbf{f}_{BB}^{k} \mathbb{E}[s_k s_k^*] \mathbf{f}_{BB}^{k\mathrm{H}} \mathbf{F}_{RF}^{\mathrm{H}} \mathbf{h}_k^* + \sum_{\substack{j=1 \\ j\neq k}}^{K} P_T \mathbf{h}_k^{\mathrm{T}} \mathbf{F}_{RF} \mathbf{f}_{BB}^{k} \mathbb{E}[s_k s_j^*] \mathbf{f}_{BB}^{j\mathrm{H}} \mathbf{F}_{RF}^{\mathrm{H}} \mathbf{h}_k^* \\
&+ \sqrt{P_T} \mathbf{h}_k^{\mathrm{T}} \mathbf{F}_{RF} \mathbf{f}_{BB}^{k} \mathbb{E}[s_k n_k^*] + \sum_{\substack{j=1 \\ j\neq k}}^{K} P_T \mathbf{h}_k^{\mathrm{T}} \mathbf{F}_{RF} \mathbf{f}_{BB}^{j} \mathbb{E}[s_j s_k^*] \mathbf{f}_{BB}^{k\mathrm{H}} \mathbf{F}_{RF}^{\mathrm{H}} \mathbf{h}_k^* \\
&+ \sum_{\substack{j=1 \\ j\neq k}}^{K} P_T \mathbf{h}_k^{\mathrm{T}} \mathbf{F}_{RF} \mathbf{f}_{BB}^{j} \mathbb{E}[s_j s_j^*] \mathbf{f}_{BB}^{j\mathrm{H}} \mathbf{F}_{RF}^{\mathrm{H}} \mathbf{h}_k^* + \sum_{\substack{j=1 \\ j\neq k}}^{K} \sqrt{P_T} \mathbf{h}_k^{\mathrm{T}} \mathbf{F}_{RF} \mathbf{f}_{BB}^{j} \mathbb{E}[s_j n_k^*] \\
&+ \sqrt{P_T} \mathbf{f}_{BB}^{k\mathrm{H}} \mathbf{F}_{RF}^{\mathrm{H}} \mathbf{h}_k^* \mathbb{E}[s_k^* n_k] + \sum_{\substack{j=1 \\ j\neq k}}^{K} \sqrt{P_T} \mathbf{f}_{BB}^{j\mathrm{H}} \mathbf{F}_{RF}^{\mathrm{H}} \mathbf{h}_k^* \mathbb{E}[s_j^* n_k] + \mathbb{E}[n_k n_k^*]
\end{aligned} \tag{C.4}
$$

assuming the cross-correlation between signal and noise, zero, which ends up with

$$
= P_T \mathbf{h}_k^{\mathrm{T}} \mathbf{F}_{RF} \mathbf{f}_{BB}^{k} \sigma_{s,k}^2 \mathbf{f}_{BB}^{k\mathrm{H}} \mathbf{F}_{RF}^{\mathrm{H}} \mathbf{h}_k^* + \sum_{\substack{j=1 \\ j\neq k}}^{K} P_T \mathbf{h}_k^{\mathrm{T}} \mathbf{F}_{RF} \mathbf{f}_{BB}^{j} \sigma_{s,j}^2 \mathbf{f}_{BB}^{j\mathrm{H}} \mathbf{F}_{RF}^{\mathrm{H}} \mathbf{h}_k^* + \sigma_n^2. \tag{C.5}
$$

The SINR_k per user would be

$$
\mathrm{SINR}_k = \frac{P_T \mathbf{h}_k^{\mathrm{T}} \mathbf{F}_{RF} \mathbf{f}_{BB}^{k} \sigma_{s,k}^2 \mathbf{f}_{BB}^{k\mathrm{H}} \mathbf{F}_{RF}^{\mathrm{H}} \mathbf{h}_k^*}{\sum\limits_{\substack{j=1 \\ j\neq k}}^{K} P_T \mathbf{h}_k^{\mathrm{T}} \mathbf{F}_{RF} \mathbf{f}_{BB}^{j} \sigma_{s,j}^2 \mathbf{f}_{BB}^{j\mathrm{H}} \mathbf{F}_{RF}^{\mathrm{H}} \mathbf{h}_k^* + \sigma_n^2}. \tag{C.6}
$$

C.2 MSE minimization for kth user

Under the independence assumption of s_k and n_k, the MSE for kth user $\mathbf{E}_k$ can be written given that $y_k = \hat{s}_k$

$$
\begin{aligned}
e_k &= \mathbb{E}\left[(\hat{s}_k - s_k) (\hat{s}_k - s_k)^{\mathrm{H}} \right] \\
&= \mathbb{E}\left[(\hat{s}_k - s_k) \left(\hat{s}_k^{\mathrm{H}} - s_k^{\mathrm{H}} \right) \right] \\
&= \mathbb{E}\left[\left(\hat{s}_k \hat{s}_k^{\mathrm{H}} - \hat{s}_k s_k^{\mathrm{H}} - s_k \hat{s}_k^{\mathrm{H}} + s_k s_k^{\mathrm{H}} \right) \right] \\
&= \mathbb{E}\left[\hat{s}_k \hat{s}_k^{\mathrm{H}} \right] - \mathbb{E}\left[\hat{s}_k s_k^{\mathrm{H}} \right] - \mathbb{E}\left[s_k \hat{s}_k^{\mathrm{H}} \right] + \mathbb{E}\left[s_k s_k^{\mathrm{H}} \right]
\end{aligned} \tag{C.7}
$$

Let us first compute the value of $\hat{s}_k^{\mathrm{H}}$ as

$$
\hat{s}_k^{\mathrm{H}} = \left(\sqrt{P_T} \mathbf{h}_k^{\mathrm{T}} \mathbf{F}_{RF} \mathbf{f}_{BB}^{k} s_k + \sum_{\substack{j=1 \\ j\neq k}}^{K} \sqrt{P_T} \mathbf{h}_k^{\mathrm{T}} \mathbf{F}_{RF} \mathbf{f}_{BB}^{j} s_j + n_k \right)^{\mathrm{H}} \tag{C.8}
$$

$$
\hat{s}_k^{\mathrm{H}} = \sqrt{P_T} s_k^{\mathrm{H}} \mathbf{f}_{BB}^{k\mathrm{H}} \mathbf{F}_{RF}^{\mathrm{H}} \mathbf{h}_k^* + \sum_{\substack{j=1 \\ j\neq k}}^{K} \sqrt{P_T} s_j^{\mathrm{H}} \mathbf{f}_{BB}^{j\mathrm{H}} \mathbf{F}_{RF}^{\mathrm{H}} \mathbf{h}_k^* + n_k^*. \tag{C.9}
$$

Computing separately each term from (C.7), we have

$$\mathbb{E}\left[\hat{s}_k \hat{s}_k^{\mathrm{H}}\right] = \mathbb{E}\left[\left(\sqrt{P_T}\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{k}s_k + \sum_{\substack{j=1\\j\neq k}}^{K}\sqrt{P_T}\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{j}s_j + n_k\right)\right.$$
$$\left.\left(\sqrt{P_T}s_k^{\mathrm{H}}\mathbf{f}_{BB}^{k\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^{*} + \sum_{\substack{j=1\\j\neq k}}^{K}\sqrt{P_T}s_j^{\mathrm{H}}\mathbf{f}_{BB}^{j\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^{*} + n_k^{*}\right)\right] \tag{C.10}$$
$$= \mathbb{E}\left[P_T\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{k}s_k s_k^{\mathrm{H}}\mathbf{f}_{BB}^{k\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^{*}\right]$$
$$+ \mathbb{E}\left[\sum_{\substack{j=1\\j\neq k}}^{K}P_T\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{j}s_j s_j^{\mathrm{H}}\mathbf{f}_{BB}^{j\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^{*}\right] + \mathbb{E}\left[n_k n_k^{*}\right] \tag{C.11}$$
$$= P_T\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{k}\sigma_{s,k}^2\mathbf{f}_{BB}^{k\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^{*} + \sum_{\substack{j=1\\j\neq k}}^{K}P_T\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{j}\sigma_{s,j}^2\mathbf{f}_{BB}^{j\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^{*} + \sigma_{n,k}^2. \tag{C.12}$$

Now calculating the other part

$$\mathbb{E}\left[\hat{s}_k s_k^{\mathrm{H}}\right] = \mathbb{E}\left[\sqrt{P_T}\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{k}s_k s_k^{\mathrm{H}} + \sum_{\substack{j=1\\j\neq k}}^{K}\sqrt{P_T}\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{j}s_j s_k^{\mathrm{H}} + n_k s_k^{\mathrm{H}}\right] \tag{C.13}$$
$$= \sqrt{P_T}\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{k}\sigma_{s,k}^2. \tag{C.14}$$

Now

$$\mathbb{E}\left[s_k \hat{s}_k^{\mathrm{H}}\right] = \sqrt{P_T}\mathbf{f}_{BB}^{k\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^{*}\sigma_{s,k}^2. \tag{C.15}$$

and finally

$$\mathbb{E}\left[s_k s_k^{\mathrm{H}}\right] = \sigma_{s,k}^2 = 1. \tag{C.16}$$

Now (C.7) can be re-written as

$$e_k = P_T\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{k}\sigma_{s,k}^2\mathbf{f}_{BB}^{k\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^{*} + \sum_{\substack{j=1\\j\neq k}}^{K}P_T\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{j}\sigma_{s,j}^2\mathbf{f}_{BB}^{j\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^{*} + \sigma_{n,k}^2$$
$$- \sqrt{P_T}\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{k}\sigma_{s,k}^2 - \sqrt{P_T}\mathbf{f}_{BB}^{k\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^{*}\sigma_{s,k}^2 + \sigma_{s,k}^2 \tag{C.17}$$

where $\sigma_{s,k}^2 = 1$ and $\sigma_{s,j}^2 = 1$

$$e_k = \left(1 - \sqrt{P_T}\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{k}\right)\left(1 - \sqrt{P_T}\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{k}\right)^{\mathrm{H}}$$
$$+ \sum_{\substack{j=1\\j\neq k}}^{K}P_T\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{j}\mathbf{f}_{BB}^{j\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^{*} + \sigma_{n,k}^2. \tag{C.18}$$

C.3 HBF-WMMSE for MU-MISO

The update of the weight variable w_k is in closed form (4.37) which is given as

$$w_k'^{\text{opt}} = e_k^{-1}. \tag{C.19}$$

The update of the transmit baseband beamforming $\mathbf{f}_{BB}^k$ for all K resulting in the following optimization problem (4.32a)

$$\arg\min_{\mathbf{f}_{BB}} \sum_{k=1}^{K} \beta_k \left(w_k' \left(\left(1 - \sqrt{P_T}\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^k\right)\left(1 - \sqrt{P_T}\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^k\right)^{\mathrm{H}} \right.\right. \tag{C.20a}$$

$$\left.\left. + \sum_{\substack{j=1 \\ j\neq k}}^{K} P_T \mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^j \mathbf{f}_{BB}^{j\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^* + \sigma_{n,k}^2 \right) - \log w_k' \right) \tag{C.20b}$$

$$\text{subject to} \quad \sum_{k=1}^{K} \mathrm{tr}\left\{\mathbf{f}_{BB}^k \mathbf{f}_{BB}^{k\mathrm{H}}\right\} \leq P_T \tag{C.20c}$$

this is a convex quadratic problem which can be solved be using standard convex optimization algorithms. In fact, this problem has a closed form solution using the Lagrangian multiplier method. Specifically, attaching the Lagrange multiplier λ to the power budget of transmitter. We get the following Lagrange function

$$\begin{aligned} L\left(\mathbf{F}_{BB}, \lambda\right) = & \sum_{k=1}^{K} \beta_k \left(w_k' \left(P_T \mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^k \mathbf{f}_{BB}^{k\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^* - \sqrt{P_T}\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^k \right.\right. \\ & \left.\left. - \sqrt{P_T}\mathbf{f}_{BB}^{k\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^* + 1 \right) \right) + \sum_{k=1}^{K}\sum_{\substack{j=1 \\ j\neq k}}^{K} \beta_k w_k' P_T \mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^j \mathbf{f}_{BB}^{j\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^* + \sigma_{n,k}^2 \\ & - \sum_{k=1}^{K} \beta_k \log w_k' + \lambda \left(\sum_{k=1}^{K} \mathrm{tr}\left\{\mathbf{f}_{BB}^k \mathbf{f}_{BB}^{k\mathrm{H}}\right\} - P_T \right), \end{aligned} \tag{C.21}$$

further solving

$$\begin{aligned} L\left(\mathbf{F}_{BB}, \lambda\right) = & \sum_{k=1}^{K} \beta_k \left(w_k' \left(1 - \sqrt{P_T}\mathbf{f}_{BB}^{k\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^* - \sqrt{P_T}\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^k \right) \right) \\ & + \sum_{k=1}^{K} \beta_k \left(w_k' P_T \mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^k \mathbf{f}_{BB}^{k\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^* \right) \\ & + \sum_{k=1}^{K}\sum_{\substack{j=1 \\ j\neq k}}^{K} \left(\beta_k w_k' P_T \mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^j \mathbf{f}_{BB}^{j\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^* + \sigma_{n,k}^2 \right) \\ & - \sum_{k=1}^{K} \beta_k \log w_k' + \lambda \left(\sum_{k=1}^{K} \mathrm{tr}\left\{\mathbf{f}_{BB}^k \mathbf{f}_{BB}^{k\mathrm{H}}\right\} - P_T \right). \end{aligned} \tag{C.22}$$

$$L\left(\mathbf{F}_{BB},\lambda\right)=\sum_{k=1}^{K}\beta_k\left(w'_k\left(1-\sqrt{P_T}\mathbf{f}_{BB}^{k\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^*-\sqrt{P_T}\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{k}\right)\right)$$
$$+\sum_{k=1}^{K}\sum_{j=1}^{K}\left(\beta_k w'_k P_T\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{j}\mathbf{f}_{BB}^{j\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^*+\sigma_{n,k}^2\right)$$
$$-\sum_{k=1}^{K}\beta_k\log w'_k+\lambda\left(\sum_{k=1}^{K}\mathrm{tr}\left\{\mathbf{f}_{BB}^{k}\mathbf{f}_{BB}^{k\mathrm{H}}\right\}-P_T\right). \tag{C.23}$$

The concentrated function is

$$L\left(\mathbf{F}_{BB},\lambda\right)=\sum_{k=1}^{K}\beta_k\left(w'_k-w'_k\sqrt{P_T}\mathbf{f}_{BB}^{k\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^*-w'_k\sqrt{P_T}\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{k}\right)$$
$$+\sum_{k=1}^{K}\sum_{j=1}^{K}\left(\beta_k w'_k P_T\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{j}\mathbf{f}_{BB}^{j\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^*\right)+\lambda\left(\sum_{k=1}^{K}\mathrm{tr}\left\{\mathbf{f}_{BB}^{k}\mathbf{f}_{BB}^{k\mathrm{H}}\right\}-P_T\right). \tag{C.24}$$

The first order optimality condition of $L\left(\left\{\mathbf{f}_{BB}^{k}\right\}_{k=1}^{K},\lambda\right)$ with respect to each $\mathbf{f}_{BB}^{l}$

$$\frac{\partial L\left(\mathbf{F}_{BB},\lambda\right)}{\partial\mathbf{f}_{BB}^{l}}=-\frac{\partial\left(\sum_{k=1}^{K}\beta_k\left(w'_k\sqrt{P_T}\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{k}\right)\right)}{\partial\mathbf{f}_{BB}^{l}}$$
$$+\frac{\partial\left(\sum_{k=1}^{K}\sum_{j=1}^{K}\left(\beta_k w'_k P_T\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{j}\mathbf{f}_{BB}^{j\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^*\right)\right)}{\partial\mathbf{f}_{BB}^{l}}+\lambda\frac{\partial\left(\sum_{k=1}^{K}\mathrm{tr}\left\{\mathbf{f}_{BB}^{k}\mathbf{f}_{BB}^{k\mathrm{H}}\right\}-P_T\right)}{\partial\mathbf{f}_{BB}^{l}}. \tag{C.25}$$

By applying derivation to individual terms [1]

$$-\frac{\partial\left(\sum_{k=1}^{K}\beta_k\left(w'_k\sqrt{P_T}\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{k}\right)\right)}{\partial\mathbf{f}_{BB}^{l}}=-\beta_l w'_l\sqrt{P_T}\mathbf{h}_l^{\mathrm{T}}\mathbf{F}_{RF} \tag{C.26}$$

$$\frac{\partial\left(\sum_{k=1}^{K}\sum_{j=1}^{K}\left(\beta_k w'_k P_T\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{j}\mathbf{f}_{BB}^{j\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^*\right)\right)}{\partial\mathbf{f}_{BB}^{l}}= \tag{C.27}$$

$$=\frac{\partial\sum_{k=1}^{K}\left[\beta_k w'_k P_T\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{1}\mathbf{f}_{BB}^{1\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^*+\cdots+\underbrace{\beta_k w'_k P_T\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{j}\mathbf{f}_{BB}^{j\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^*}_{j=l}+\ldots\right.}{\partial\mathbf{f}_{BB}^{l}}$$
$$\left.\frac{+\beta_k w'_k P_T\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF}\mathbf{f}_{BB}^{K}\mathbf{f}_{BB}^{K\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^*\right]}{\partial\mathbf{f}_{BB}^{l}} \tag{C.28}$$

as derivaive is linear operator, so we can apply derivation inside on each term individually

$$=\sum_{k=1}^{K}\beta_k w'_k P_T\mathbf{f}_{BB}^{l\mathrm{H}}\mathbf{F}_{RF}^{\mathrm{H}}\mathbf{h}_k^*\mathbf{h}_k^{\mathrm{T}}\mathbf{F}_{RF} \tag{C.29}$$

[1] $\frac{\partial\left(\mathbf{a}^{\mathrm{T}}\mathbf{X}\mathbf{b}\right)}{\partial\mathbf{X}}=\mathbf{b}\mathbf{a}^{\mathrm{T}}$

and finally [2]

$$\lambda \frac{\partial \left(\sum_{k=1}^{K} \mathrm{tr} \left\{ \mathbf{f}_{BB}^{k} \mathbf{f}_{BB}^{k\mathrm{H}} \right\} - P_T \right)}{\partial \mathbf{f}_{BB}^{l}} = \lambda \mathbf{f}_{BB}^{l\mathrm{H}} \tag{C.30}$$

which turns out to be

$$-\beta_l w_l' \sqrt{P_T} \mathbf{h}_l^{\mathrm{T}} \mathbf{F}_{RF} + \sum_{k=1}^{K} \beta_k w_k' P_T \mathbf{f}_{BB}^{l\mathrm{H}} \mathbf{F}_{RF}^{\mathrm{H}} \mathbf{h}_k^{*} \mathbf{h}_k^{\mathrm{T}} \mathbf{F}_{RF} + \lambda \mathbf{f}_{BB}^{l\mathrm{H}} = 0 \tag{C.31}$$

$$\mathbf{f}_{BB}^{l\mathrm{H}} \left[\sum_{k=1}^{K} \beta_k w_k' P_T \mathbf{F}_{RF}^{\mathrm{H}} \mathbf{h}_k^{*} \mathbf{h}_k^{\mathrm{T}} \mathbf{F}_{RF} + \lambda \mathbf{I}_K \right] = \beta_l w_l' \sqrt{P_T} \mathbf{h}_l^{\mathrm{T}} \mathbf{F}_{RF} \tag{C.32}$$

$$\mathbf{f}_{BB}^{l\mathrm{H}} = \beta_l w_l' \sqrt{P_T} \mathbf{h}_l^{\mathrm{T}} \mathbf{F}_{RF} \left[\sum_{k=1}^{K} \beta_k w_k' P_T \mathbf{F}_{RF}^{\mathrm{H}} \mathbf{h}_k^{*} \mathbf{h}_k^{\mathrm{T}} \mathbf{F}_{RF} + \lambda \mathbf{I}_K \right]^{-1} \tag{C.33}$$

$$\mathbf{f}_{BB}^{l\mathrm{opt}} = \left[\sum_{k=1}^{K} \beta_k w_k' P_T \mathbf{F}_{RF}^{\mathrm{H}} \mathbf{h}_k^{*} \mathbf{h}_k^{\mathrm{T}} \mathbf{F}_{RF} + \lambda \mathbf{I}_K \right]^{-1} \beta_l w_l' \sqrt{P_T} \mathbf{F}_{RF}^{\mathrm{H}} \mathbf{h}_l^{*}, \quad l = 1, \ldots, K. \tag{C.34}$$

[2] $\frac{\partial \mathrm{tr}\left\{ \mathbf{X}^{\mathrm{T}} \mathbf{A} \right\}}{\partial \mathbf{X}} = \mathbf{A}$ and $\partial \left(\mathrm{tr} \left\{ \mathbf{X} \right\} = \mathrm{tr} \left\{ \partial \left(\mathbf{X} \right) \right\} \right)$