KU LEUVEN

ARENBERG DOCTORAL SCHOOL
Faculty of Engineering Technology

Learning to Precode for Energy Reduction in Massive MIMO Systems under Hardware Nonlinearities

Thomas Feys

Dissertation presented in partial fulfillment
of the requirements for the degree of
Doctor of Engineering Technology (PhD)

January 2026

Supervisors:
Prof. dr. ir. F. Rottenberg
Prof. dr. ir. L. Van der Perre
Prof. dr. ir. G. Callebaut

LEARNING TO PRECODE FOR ENERGY REDUCTION IN MASSIVE MIMO SYSTEMS UNDER HARDWARE NONLINEARITIES

Thomas FEYS

Supervisors:
Prof. dr. ir. F. Rottenberg
Prof. dr. ir. L. Van der Perre
Prof. dr. ir. G. Callebaut

Members of the
Examination Committee:
Prof. dr. ir. E. Gruyaert, chair
Prof. dr. ir. A. Bertrand
Prof. dr. ir. S. Pollin
Prof. dr. ir. T. Eriksson
(Chalmers University of Technology, Sweden)
ir. B. Sihlbom
(Huawei, Sweden)

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Engineering Technology (PhD)

January 2026

© 2026 Thomas Feys
Uitgegeven in eigen beheer, Thomas Feys, Gebroeders De Smetstraat 1, B-9000 Ghent (Belgium)

Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt worden door middel van druk, fotokopie, microfilm, elektronisch of op welke andere wijze ook zonder voorafgaande schriftelijke toestemming van de uitgever.

All rights reserved. No part of the publication may be reproduced in any form by print, photoprint, microfilm, electronic or any other means without written permission from the publisher.

Preface

> "It's only a point [...] When you're playing a point, it has to be the most important thing in the world and it is. But when it's behind you, it's behind you."
>
> Roger Federer

The PhD journey can often feel like a long rally: countless experiments, revisions, and moments of uncertainty. Like a tennis match, progress in research isn't about perfection but persistence, showing up for the next point, the next idea, the next draft. In this sense, the work contained here represents not a conclusion but a continuum of learning and effort. Similar to a tennis match, a PhD can appear to be a 'solo sport'. However, the support of a team on the sidelines is invaluable. It is therefore only fitting to thank those who have made this journey possible.

First of all, I would like to thank my supervisor, François, for his constant support throughout this PhD. I am especially grateful for his availability and for the time he always took to discuss technical and mathematical details. After finishing my master's degree in Artificial Intelligence, I thought I had a solid grasp of how machine learning algorithms work. Being asked to explain their mathematical foundations pushed me to look deeper, question my assumptions, and truly understand what was going on under the hood. This has been invaluable to my development as a researcher. I am also very thankful for the opportunity to do a research stay in Barcelona. Beyond the academic experience, this stay introduced me to amazing people and turned Barcelona into a place that now feels like a second home, one where I always have friends to meet whenever I return.

I would also like to thank Liesbet for giving me the opportunity to pursue a PhD and for introducing me to François. Despite having an extremely busy schedule, she always made time when it truly mattered. Her feedback consistently encouraged me to step back and look at the bigger picture, which has been incredibly valuable.

I am equally grateful to Gilles for his constant availability and willingness to help. Beyond the occasional random yelling at his computer, it has been truly insightful and fun, to work with him. Many challenging moments were brightened by his input, enthusiasm and by partaking in the weekly fatty Friday ritual.

I would like to express my gratitude to my supervisory committee, Prof. A. Bertrand and Ir. B. Sihlbom, for their guidance, insightful feedback, and support throughout my PhD journey. I am also grateful to the examination committee, Prof. S. Pollin and Prof. T. Eriksson, for taking the time to review this work and provide valuable comments that helped strengthen the thesis. Finally, I would like to thank Prof. E. Gruyaert for chairing the defense.

Besides my supervisors, the most vital part of my team has undoubtedly been my Dramco colleagues. When I started this PhD, I vividly remember the strange feeling of walking into an office filled with PhD students who had previously tutored the countless lab sessions during my master's program. One might expect this to feel uncomfortable, and for the first few weeks, it did. However, the Dramco team quickly grew into a family, fostering a truly unbeatable atmosphere. I would also like to thank several colleagues who made the day-to-day of this PhD so enjoyable. Special thanks to Tijl for carrying much of the work on the 6GTandem simulator and for gracefully diffusing every new "idea" that Gilles threw our way. I am grateful to Guus for always being ready to discuss renovation plans and for bringing his dog to brighten up the office. Many thanks to Daan and Geoffrey for being equally excited for the Friday (and occasional weekday) pitta ritual. I am also thankful to Bert C., Jona and Chesney for suffering through the Kom op tegen Kanker 100 km runs together. A special thanks goes to Chesney for always being gracious in defeat during our ping-pong matches. Finally, I want to thank Emanuele for joining me as one of the first PhD students under François. From ping-pong and tennis matches to watching Grand Slam finals together in the office, these moments created memories I will always cherish. I would also like to thank Jona, Jorik, Bert P., Tijl and Daan for the many early-morning discussions. To Alex, Fan, Jarne, Jean Pierre, Lieven, Lyssa, Nobby, Reza, Sam, Sarah, Sogand, Tianzheng, Willem and Youssef thank you for the great Dramco atmosphere.

I would also like to thank my friends and family for their unwavering support throughout this journey. Special thanks to my parents, who have always encouraged me. Their guidance has shaped me into the person I am today. A special thanks goes to Stephan for always being available to play tennis, providing a welcome distraction and a way to take my mind of the PhD. To my friends, thank you for the laughter, companionship, and support that kept life balanced beyond the lab. Your encouragement, patience, and presence have made the most challenging days manageable and the joyful moments even sweeter.

Finally, I would like to thank Cathy. Doing a PhD can feel isolating, as few people truly understand the process and the struggles it brings. The wonderful coincidence that we both decided to pursue a PhD made the journey much more enjoyable. Even though we work in different fields, the challenges of a PhD are remarkably similar. I couldn't have asked for anyone better to share the ups and downs with, to celebrate accepted papers, to complain about reviewer two's comments, and to simply talk through the highs and lows.

Thank you for always supporting me and for motivating me to do the administrative tasks I so dislike. But most importantly, thank you for not ending every argument over the last few months with "Who is the doctor in this household?". After completing this thesis, finally, we are equal again.

Popularized Abstract

Future wireless networks aim for higher data rates, better coverage, and lower latency, however, there is a hidden cost behind every video call, stream, and cloud service: a significant environmental footprint. Today, mobile networks already account for a significant share of global electricity use and carbon emissions, with base stations being the main consumers. As traffic keeps growing and 5G and future 6G systems deploy more antennas and wider bandwidths, simply “adding more hardware” has turned our communication infrastructure into an energy-hungry system.

This PhD thesis asks a simple but crucial question: how can base stations use far less energy without sacrificing the quality and speed that users expect? Modern base stations rely on a technique called massive multiple input multiple output (MIMO). It uses many antennas to beam signals directly to multiple users at once. This provides excellent coverage and high data rates, but it also means a lot of radio hardware that needs power, cooling, and complex signal processing. This work focuses on two components that hold a significant share of this energy budget: the power amplifiers that boost the signals before they are transmitted over the air, and the digital-to-analog converters that translate digital bits into analog signals. These components are usually operated very conservatively to keep the signal "clean", which is useful from a performance and compliance perspective, but comes at a considerable energy cost. More precisely, power amplifiers are held back from their most efficient operating region to avoid distorting the signal, and digital-to-analog converters use many bits of precision to keep quantization errors low. These design choices improve signal quality, but they also mean that a lot of potential energy savings are left on the table.

This thesis explores a different philosophy: instead of insisting that every hardware block behaves perfectly, misbehavior is allowed on purpose. In practical terms, this means driving power amplifiers closer to saturation, where they are much more efficient but become nonlinear and distort the signal, and reducing the number

of bits in digital-to-analog converters, which saves power but introduces coarse quantization noise. On paper, both steps degrade the signal quality. However, in this thesis, the central idea is to allow these imperfections and then compensate for them in software.

The key enabler is a class of machine learning models called neural networks, and more specifically, graph neural network (GNN), which are particularly well suited for systems with many interacting elements, such as antenna arrays that serve multiple users. In a massive MIMO base station, dozens of antennas cooperate to beam signals towards several users at once, and the "precoder" is the algorithm that decides how each antenna should adjust the amplitude and phase of its signal. By redesigning this precoder as a learned GNN-based function, this work shows that the base station can anticipate the distortions created by imperfect hardware and shape the transmitted signal so that these distortions are pushed away from the users, as much as possible.

At the same time, this thesis emphasizes that such learning-based precoders are not a universal solution and come with their own computational and energy costs. For each proposed neural network architecture, the additional processing power required to run the neural network is quantified and compared to the savings achieved in the power amplifiers and digital-to-analog converters. With respect to the power amplifiers, the results show that, with the right model structure, learning to precode with realistic, imperfect hardware can yield a net reduction in energy consumption. For digital-to-analog converters, the trade-off is more nuanced as the neural network complexity is high, resulting in limited gains that only hold for low bandwidth systems.

Overall, this thesis shows that "learning to precode" with realistic, imperfect hardware can be a powerful lever to make future wireless networks more energy-efficient, provided that the trade-off between hardware savings and algorithmic complexity is explicitly accounted for. The proposed graph neural network architectures, together with the respective power-consumption analyses, provide practical building blocks for such distortion-aware transmitters and illustrate how machine learning can contribute to, but does not automatically guarantee, both economic and ecological improvements in next-generation communication systems.

Gepopulariseerde Samenvatting

Toekomstige draadloze netwerken streven naar hogere datasnelheden, betere dekking en lagere vertraging, maar achter elk videogesprek, elke stream en elke clouddienst schuilt een verborgen prijs: een aanzienlijke impact op het milieu. Tegenwoordig zijn mobiele netwerken al verantwoordelijk voor een groot deel van het wereldwijde elektriciteitsverbruik en de CO2-uitstoot, waarbij basisstations de belangrijkste verbruikers zijn. Naarmate het dataverkeer blijft groeien en 5G- en toekomstige 6G-systemen meer antennes en bredere bandbreedtes inzetten, blijkt "gewoon meer hardware toevoegen" er voor te zorgen dat onze communicatiesystemen veel energie verbruiken.

Deze thesis stelt een eenvoudige maar cruciale vraag: hoe kunnen basisstations minder energie gebruiken zonder in te leveren op de kwaliteit en snelheid die gebruikers verwachten? Moderne basisstations maken gebruik van een techniek genaamd massive multiple input multiple output (MIMO). Hierbij worden veel antennes gebruikt om gerichte signalen tegelijk naar meerdere gebruikers te sturen. Dit biedt uitstekende dekking en hoge datasnelheden, maar het betekent ook veel radiohardware die energie, koeling en complexe signaalverwerking nodig heeft. Deze thesis richt zich op twee componenten die een aanzienlijk deel van dit energieverbruik vertegenwoordigen: de versterkers die de signalen versterken voordat ze worden uitgezonden, en de digitaal-naar-analoog convertoren die digitale bits vertalen naar analoge signalen. Deze componenten worden meestal zeer conservatief gebruikt om het signaal "schoon" te houden, wat goed is voor prestaties en naleving van normen, maar dit gaat ten koste van veel energie. Concreet worden versterkers beperkt tot minder efficiënte werkgebieden om vervorming van het signaal te voorkomen, en digitaal-naar-analoog convertoren gebruiken veel bits om kwantisatiefouten laag te houden. Deze ontwerpkeuzes verbeteren de signaalkwaliteit, maar laten tegelijkertijd veel potentiële energiebesparingen

onbenut.

Deze thesis verkent een andere aanpak: in plaats van te eisen dat elk hardwareblok perfect werkt, wordt opzettelijk ruimte gelaten voor imperfectie. Praktisch betekent dit dat versterkers dichter bij verzadiging worden gestuurd, waar ze veel efficiënter zijn maar het signaal vervormen, en dat het aantal bits in digitaal-naar-analoog convertoren wordt verminderd, wat energie bespaart maar kwantisatieruis introduceert. Op papier verslechtert dit de signaalkwaliteit. In dit werk is het centrale idee echter om deze imperfecties toe te staan en ze vervolgens in software te compenseren.

De sleutel tot deze aanpak is een klasse van machine learning modellen genaamd neurale netwerken, en meer specifiek graaf neurale netwerken, die bijzonder geschikt zijn voor systemen met veel onderling interagerende elementen, zoals antennearrays die meerdere gebruikers dekken. In een massive MIMO basisstation werken tientallen antennes samen om signalen tegelijk naar meerdere gebruikers te sturen, en de "precoder" is het algoritme dat bepaalt hoe de amplitude en fase van het signaal aan elke antenne aangepast moet worden. Door deze precoder te ontwerpen als een graaf neuraal netwerk, toont dit onderzoek aan dat vervormingen door imperfecte hardware voorspeld kunnen worden en weggestuurd kunnen worden van de gebruikers.

Tegelijkertijd benadrukt deze thesis dat dergelijke neurale netwerken geen universele oplossing zijn en hun eigen reken- en energieverbruik met zich meebrengen. Voor elke voorgestelde neurale netwerkarchitectuur wordt het extra energieverbruik dat nodig is om het netwerk te gebruiken gekwantificeerd en vergeleken met de energie besparingen bij de versterkers en digitaal-naar-analoog convertoren. Voor versterkers tonen de resultaten aan dat, met de juiste modelstructuur, leren precoderen met realistische, imperfecte hardware kan leiden tot een netto verlaging van het energieverbruik. Voor digitaal-naar-analoog convertoren is de afweging genuanceerder: de complexiteit van het neurale netwerk is hoog, waardoor de winst beperkt blijft en alleen geldt voor systemen met lage bandbreedte.

Als besluit laat deze thesis zien dat leren precoderen met realistische, imperfecte hardware een krachtig instrument kan zijn om toekomstige draadloze netwerken energiezuiniger te maken, mits de afweging tussen hardwarebesparingen en algoritmische complexiteit expliciet wordt meegenomen. De voorgestelde graaf neurale netwerkarchitecturen, samen met de bijbehorende analyses van het energieverbruik, bieden praktische bouwstenen voor dergelijke vervormingsbewuste zenders en illustreren hoe machine learning kan bijdragen tot, maar niet automatisch resulteert in, economische en ecologische verbeteringen van communicatiesystemen.

Abstract

The main goal of this thesis is to reduce the energy consumption of massive multiple input multiple output (MIMO) systems, while maintaining the quality of service the user expects. This goal is addressed by allowing hardware components in the base station that consume a lot of power to operate in more energy-efficient regimes. Traditionally, these components are operated conservatively to avoid signal degradation. However, this safety margin results in a lot of potential energy savings being left on the table. Concretely, this work aims at reducing the energy consumption of power amplifiers and digital-to-analog converters (DACs). Operating the power amplifiers in an energy-efficient regime introduces nonlinear distortion, which deteriorates the signal quality. Similarly, DACs consume less power if they use fewer bits. However, using fewer bits results in quantization-induced distortion, which, again, deteriorates the signal quality. These distortions are traditionally avoided given that they lead to complex optimization problems which are non-convex and include non-differentiable and non-continuous functions. Nevertheless, in this work, these distortions are accepted and overcome by developing neural network (NN)-based precoders that mitigate the distortion.

For power amplifier distortion, a linear precoder is proposed that maximizes the sum rate in the presence of nonlinear distortion. A graph neural network (GNN) is proposed that leverages the natural permutation equivariance of the precoding problem with respect to the antenna and user dimensions. This provides a hypothesis space that is greatly reduced, resulting in a NN with fewer trainable parameters that is easier to train. This GNN-based distortion-aware precoder is developed both for the single-carrier and multi-carrier cases. In both cases, self-supervised training is used to directly maximize the achievable sum rate under nonlinear power amplifiers. This embeds the underlying physics and hardware behavior directly into the loss function and avoids the need for labeled optimal precoders that are hard or impossible to compute. By using such NN-based

precoders, distortion is transmitted in non-user directions. This delivers significant energy savings for a fixed quality of service constraint. Moreover, it is shown that the energy savings are not outweighed by the additional power consumption required for GNN inference.

When looking at DACs, GNN-based nonlinear symbol-level precoding is proposed to mitigate quantization-induced distortion. This nonlinear precoding problem exhibits permutation equivariance with respect to the antenna dimension and permutation invariance with respect to the user dimension. These properties are enforced in the GNN architecture. Similar to the linear precoding case, this NN architecture results in a greatly reduced hypothesis space, which produces NNs with fewer trainable parameters, which are easier to train. Given the NP-hard nature of the quantized non-linear precoding problem, obtaining optimal ground truth labels is computationally expensive. To overcome this, a self-supervised loss function is proposed, which directly maximizes the achievable rate. However, the computation of the achievable rate requires the use of the non-differentiable and non-continuous DAC functions. To mitigate this, a Gumbel-softmax-based relaxation is introduced to estimate the gradients of the loss function. It is shown that the combination of this GNN architecture and loss function produces significant performance gains. However, overall energy savings are limited in this case due to the significant inference complexity stemming from the increased model size and the symbol-level inference rate.

Overall, this work demonstrates that neural network-based precoders, and in particular problem-specific GNN architectures trained with self-supervised, physics-informed loss functions, can enable meaningful energy reductions in massive MIMO systems by allowing certain hardware blocks to operate in more efficient regimes while managing the resulting distortion. At the same time, the results underline that such learning-based schemes are not universally beneficial: their advantages depend critically on the operating point, the choice of NN architecture, and the balance between hardware power savings and additional digital processing power. The most pronounced net gains are obtained for linear precoding targeting power amplifiers, while for DACs, the trade-off is more nuanced as the neural network complexity is high, resulting in limited gains that only hold for low bandwidth systems.

Beknopte Samenvatting

Het hoofddoel van deze thesis is het verminderen van het energieverbruik van massive multiple input multiple output (MIMO)-systemen, terwijl de door de gebruiker verwachte kwaliteit van dienstverlening behouden blijft. Dit doel wordt nagestreefd door toe te staan dat energie-intensieve hardwarecomponenten in het basisstation in energie-efficiëntere werkgebieden opereren. Traditioneel worden deze componenten zeer conservatief aangestuurd om signaaldegradatie te vermijden. Deze veiligheidsmarges zorgen er echter voor dat veel potentiële energiebesparingen onbenut blijven. Concreet richt dit werk zich op het verlagen van het energieverbruik van versterkers en digitaal-naar-analoog convertoren. Het aansturen van versterkers in een energie-efficiënter regime introduceert niet-lineaire vervorming, wat de signaalkwaliteit verslechtert. Op vergelijkbare wijze verbruiken digitaal-naar-analoog convertoren minder energie wanneer ze met minder bits werken, maar dit leidt tot kwantisatie fouten, die eveneens de signaalkwaliteit aantasten. Dergelijke vervormingen worden traditioneel vermeden omdat zij leiden tot complexe optimalisatieproblemen die niet-convex zijn en niet-differentieerbare en niet-continue functies bevatten. In deze thesis worden deze vervormingen geaccepteerd en gecompenseerd door de ontwikkeling van precoders die gebaseerd zijn op neurale netwerken.

Voor versterkers die distorsie genereren wordt een lineaire precoder voorgesteld die de datasnelheid maximaliseert in aanwezigheid van niet-lineaire vervorming. Hiervoor wordt een graaf neuraal netwerk geïntroduceerd die gebruikmaakt van de permutatie-equivariantie van het precoderings probleem ten opzichte van de antenne- en gebruikersdimensies. Dit resulteert in een sterk gereduceerde hypotheseruimte, wat leidt tot een neuraal netwerk met minder trainbare parameters die eenvoudiger te trainen is. Deze vervormingsbewuste precoder wordt ontwikkeld voor zowel single-carrier- als multi-carriersystemen. In beide gevallen wordt self-supervised training toegepast om de sumrate onder niet-lineaire versterkers rechtstreeks te maximaliseren. Hierdoor wordt de onderliggende fysica

en het hardwaregedrag direct in de loss-functie geïntegreerd. Dit zorgt ervoor dat labels in de vorm van optimale precoders, die moeilijk of onmogelijk te berekenen zijn, niet nodig zijn. Door het gebruik van dergelijke neurale netwerk precoders wordt de vervorming weggestuurd van de gebruikers. Dit resulteert in aanzienlijke energiebesparingen voor een gelijke datasnelheid. Bovendien wordt aangetoond dat deze energiebesparingen niet teniet worden gedaan door het extra energieverbruik dat nodig is voor het neurale netwerk.

Voor digitaal-naar-analoog convertoren wordt een graaf neuraal netwerk voorgesteld voor niet-lineaire symbol-level precodering die kwantisatie ruis compenseert. Dit niet-lineaire precoderingsprobleem vertoont permutatie-equivariantie ten opzichte van de antennedimensie en permutatie-invariantie ten opzichte van de gebruikersdimensie. Deze eigenschappen worden expliciet afgedwongen door het graaf neurale netwerk. Net als in het geval van lineaire precodering resulteert dit neuraal netwerk in een sterk gereduceerde hypotheseruimte, wat leidt tot neurale netwerken met minder trainbare parameters die eenvoudiger te trainen zijn. Het berekenen van optimale precoder labels is computationeel zeer kostbaar. Om dit te omzeilen wordt een self-supervised loss-functie voorgesteld die de haalbare datasnelheid rechtstreeks maximaliseert. De berekening van deze datasnelheid vereist echter het gebruik van niet-differentieerbare en niet-continue functies. Om dit probleem te omzeilen wordt een Gumbel-softmax gebaseerde relaxatie geïntroduceerd om de gradiënten van de loss-functie te benaderen. Er wordt aangetoond dat de combinatie van deze loss-functie en het voorgestelde graaf neurale netwerk leidt tot aanzienlijke prestatieverbeteringen. Desondanks blijven de totale energiebesparingen in dit geval beperkt, vanwege de aanzienlijke complexiteit van het neurale netwerk als gevolg van de grotere modelomvang en de symbol-level berekeningstijd.

Samenvattend toont dit werk aan dat neurale netwerk precoders, en in het bijzonder probleemspecifieke graaf neurale netwerken, die worden getraind met self-supervised, fysica-geïnformeerde loss-functies, betekenisvolle energiereducties in massive MIMO systemen mogelijk maken. De energie besparing wordt verwezenlijkt door hardwareblokken in efficiëntere werkgebieden te laten werken en de resulterende distorsies te beheersen. Tegelijkertijd onderstrepen de resultaten dat dergelijke methoden niet universeel voordelig zijn: hun effectiviteit hangt sterk af van het werkregime, de keuze van neurale netwerk en de balans tussen hardware energiebesparingen en extra energie verbruik door het neurale netwerk. De meest uitgesproken netto winst wordt behaald bij lineaire precoding gericht op versterkers, terwijl de afweging voor digitaal-naar-analoog convertoren genuanceerder is vanwege de hoge complexiteit van het neurale netwerk, wat resulteert in beperkte winst die enkel geldt voor systemen met lage bandbreedte.

List of Abbreviations

3GPP third generation partnership project. 99, 148, 150

ACLR adjacent channel leakage ratio. xxvii, 19, 107, 108, 110, 143, 150

ADC analog-to-digital converter. 2, 6, 113–115, 119

AFE analog front-end. 6

AI artificial intelligence. 2, 7, 11, 145, 147–152

AM/AM amplitude modulation to amplitude modulation. xxv, 18, 31–33, 78

AM/PM amplitude modulation to phase modulation. xxv, 18, 31–33, 78, 79

AQNM additive quantization noise model. 118

AWGN additive white Gaussian noise. 77, 79

BOM bill of materials. 28, 146

BS base station. 1–7, 16, 30, 43, 58, 65, 66, 76, 85, 114, 117–119, 129, 143, 146

CCNN circular convolutional neural network. xxv, 27, 40–42

CNN convolutional neural network. 41, 115

DAB distortion-aware beamforming. xxvi, xxix, 27, 28, 37, 38, 46, 48, 49, 54, 55, 63

DAC digital-to-analog converter. ix, x, xxviii, 1, 2, 4–11, 111–116, 118, 119, 122, 126, 134–136, 138–144, 146, 147, 150, 155

DBB digital baseband. 6

DFT discrete Fourier transform. 76, 78, 83, 93, 102, 104

DPD digital pre-distortion. 15, 16, 27–29, 32, 33, 52, 55, 56, 59, 60, 63

DSP digital signal processing. 60, 63, 98, 109

eCDF empirical cumulative distribution function. xxv, 21, 23, 24

EMC electromagnetic compatibility. 150

ETSI European telecommunications standards institute. 100, 148, 150

FLOPs floating point operations. 46–49, 60, 83, 84, 87, 131, 140, 141

FR frequency range. 149

GDPR general data protection regulation. 151

GFLOPs/s giga floating point operations per second. 60, 87

GNN graph neural network. vi, ix, x, xxv–xxix, 26–28, 41–47, 49–61, 63–66, 68, 69, 72–76, 80–84, 87–93, 97, 98, 100, 102–104, 106–112, 115, 116, 128–136, 138–144, 147, 151, 152

HHI Herfindahl–Hirschman index. xxvi, 86, 91–93, 95

i.i.d. independently and identically distributed. xxvi, xxvii, 31, 65, 66, 68, 77, 88, 89, 103, 104, 110, 115, 157

IB in-band. 19

IBO input back-off. xxvi, 32, 42, 50, 55–57, 59, 63, 78, 79, 83, 88–92, 94–97, 101, 109

ICT information and communication technologies. 5, 145

IDFT inverse discrete Fourier transform. 76, 82, 83, 101, 104

ITU international telecommunication union. 148

KPI key performance indicator. 149

LoS line-of-sight. xxvi, xxvii, 14, 33, 51, 54, 57, 58, 65, 67, 69–72, 108, 121, 134, 136

LReLU leaky rectified linear unit. 51, 134

MIMO multiple input multiple output. v–xii, xxv, 1, 3–5, 7, 11, 14–16, 24, 26, 28, 30, 63, 65, 68, 74, 113–115, 143, 145, 153, 156, 157

ML machine learning. 1, 2, 7, 11, 64, 145, 147–151

MLOps machine learning operations. 152

MLP multilayer perceptron. xxv, 27, 29, 38, 40–42, 115

MMSE minimum mean squared error. 114

MMSQE minimum mean squared quantization error. 119

MRT maximum ratio transmission. xxv–xxvii, 16–24, 37, 51–54, 61, 76, 83, 84, 87–93, 96, 97, 100, 103–110, 117, 118, 122, 134–136, 138–140

NCO neural combinatorial optimization. 116

NLoS non-line-of-sight. 67

NMSE normalized mean squared error. xxix, 135, 136

NMSQE normalized mean squared quantization error. 118

NN neural network. ix, x, 7–9, 11, 27, 35–43, 47, 49, 50, 64, 65, 68, 75, 80, 89, 111, 112, 114–116, 120, 122, 123, 125–127, 129, 134, 140, 141, 143, 151–157

OFDM orthogonal frequency division multiplexing. 15, 30, 63, 100, 107, 109, 119, 149

OOB out-of-band. 14, 19, 100, 102, 103, 105, 106, 108–110

PA power amplifier. xxv, xxvi, xxix, 1, 2, 4–11, 13–18, 24–38, 42, 46, 50–52, 54–63, 65, 70, 74–76, 78–80, 83, 85, 86, 88, 90–103, 105–107, 109, 111, 113, 114, 122, 143, 146, 147, 149, 150, 153–156

PAPR peak-to-average power ratio. 15, 29

PDP power delay profile. 78

PSC power supply and cooling systems. 6

PSD power spectral density. xxvii, 106, 107, 110

QoS quality of service. 5, 7, 14, 95, 96, 109, 147, 148

RAN radio access network. 148

RF radio frequency. 6, 112, 113

RF-DAC radio frequency digital-to-analog converter. xxviii, 113, 139–142, 144

SDG Sustainable Development Goal. 145

SDR signal-to-distortion ratio. xxv–xxvii, 14, 17, 19, 34, 53, 54, 56–58, 113, 121–123, 135, 136

SNDR signal-to-noise-and-distortion ratio. 20, 21, 79, 102

SNIDR signal-to-noise-and-interference-and-distortion ratio. 20, 23, 34–36, 50, 121

SNR signal-to-noise ratio. xxv, xxvi, 4, 15–17, 28, 44, 45, 61, 63, 65, 66, 69, 96, 103, 133, 134, 139, 144

SP single precision. 61

TDL tapped delay line. 77, 88

ULA uniform linear array. 34, 54, 57, 58, 67, 107, 121, 136

WMMSE weighted minimum mean squared error. 65

Z3RO zero third-order distortion. xxv, 16–25, 30, 37, 51–53, 61

ZF zero forcing. xxvi, xxvii, xxix, 27, 37, 42, 47, 49, 54–63, 68–72, 100, 114, 118, 122, 123, 138, 139

List of Symbols and Notations

Notation:

a	Scalars are denoted using non-bold lowercase letters
$\boldsymbol{a}$	Vectors are denoted using bold lowercase letters
$\boldsymbol{A}$	Matrices are denoted using bold uppercase letters
$[\boldsymbol{A}]_{i,j}$	The element at location (i, j) in matrix $\boldsymbol{A}$
$[\boldsymbol{a}]_i$	The element at location i in vector $\boldsymbol{a}$
a_i	The element at location i in vector $\boldsymbol{a}$
$a(\cdot)$	A scalar function is denoted by a lowercase letter
$\boldsymbol{a}(\cdot)$	A vector function is denoted by a bold lowercase letter
$\boldsymbol{C}_x$	Covariance matrix of random vector $\boldsymbol{x}$
$(\cdot)^\intercal$	Transpose operator
$(\cdot)^*$	Conjugate operator
$(\cdot)^H$	Hermitian transpose operator
$\mathrm{tr}\,()$	Trace of a matrix
$\odot$	The element-wise or Hadamard product
$\mathbb{E}(.)$	Expectation
$\|\cdot\|$	L2-norm

$\mathcal{F}(\cdot)$ Discrete Fourier transform operator

$\mathcal{F}^{-1}(\cdot)$ Inverse discrete Fourier transform operator

$\Re\{x\}$ Real part of the complex number x

$\Im\{x\}$ Imaginary part of the complex number x

$\boldsymbol{I}_M$ $M \times M$ identity matrix

$\text{diag}(\boldsymbol{A})$ Main diagonal of square matrix $\boldsymbol{A}$

$\mathbb{1}_i^N$ The indicator function denoting a one-hot encoded vector of length N containing all zeros except for a one at index i

$\mathcal{G} = (\mathcal{V}, \mathcal{E})$ A graph where $\mathcal{V}$ is the set of nodes and $\mathcal{E}$ the set of edges

(a, b) The edge of a graph going from node $a \in \mathcal{V}$ to node $b \in \mathcal{V}$

$\mathcal{N}(a) = \{b \in \mathcal{V} : (a, b) \in \mathcal{E}\}$ The neighborhood of node a

Symbols:

$\boldsymbol{\Pi}$ Permutation matrix

$\mathcal{Q}(\cdot)$ Element-wise non-linear quantization function

$\boldsymbol{\phi}(\cdot)$ Element-wise non-linear power amplifier function

λ_c Wavelength at the carrier frequency

$\mathcal{CN}(0, 1)$ Complex normal distribution with zero mean and variance one

$\boldsymbol{H}$ Channel matrix

$\boldsymbol{W}$ Precoding matrix

$\boldsymbol{r}$ Received signal vector

$\boldsymbol{s}$ Vector of intended transmit symbols for all users

$\boldsymbol{v}$ Additive white Gaussian noise

$\boldsymbol{x}$ Precoded symbol vector

$\boldsymbol{y}$ Precoded symbol vector after applying a non-linear distortion function

f_c Carrier frequency

K Number of users

M	Number of transmit antennas
p_{sat}	Saturation power of the power amplifier
Q	Number of sub-carriers
R	Achievable rate

Contents

Popularized Abstract **v**

Gepopulariseerde Samenvatting **vii**

Abstract **ix**

Beknopte Samenvatting **xi**

List of Abbreviations **xvi**

List of Symbols and Notations **xix**

Contents **xxi**

List of Figures **xxv**

List of Tables **xxix**

1 Introduction **1**
1.1 From Maximum Performance to Power-Aware Design 1
1.2 Terminology and Preliminaries 3
1.2.1 Massive MIMO: a Blessing or a Curse? 3
1.2.2 Energy Consumption of Wireless Networks 5
1.2.3 Opportunities of Artificial Intelligence and Machine Learning 7
1.3 Research Questions 7
1.4 Thesis Outline and Contributions 9

2 Power Amplifier Nonlinearities & Analytical Solutions **13**
2.1 Introduction 14
2.1.1 Problem Statement 14
2.1.2 Conventional Solutions and their Limitations 15

2.1.3 Contributions . . . 16
2.2 Transmission Model . . . 16
2.2.1 Transmit Signal . . . 16
2.2.2 Z3RO and MRT Precoders . . . 17
2.2.3 PA Model . . . 18
2.2.4 Channel Measurements . . . 18
2.2.5 Received Signal . . . 19
2.3 Performance Assessment . . . 19
2.3.1 Performance Evaluation Metrics . . . 19
2.3.2 Single-User Case . . . 20
2.3.3 Two-User Case . . . 23
2.4 Conclusion . . . 24

3 Neural Network Approaches for Distortion-Aware Precoding 26
3.1 Introduction . . . 28
3.1.1 Problem Statement . . . 28
3.1.2 State-of-the-Art . . . 28
3.2 Problem Formulation . . . 30
3.2.1 System Model . . . 30
3.2.2 Non-Linear PA Models . . . 31
3.2.3 Radiation Patterns . . . 33
3.2.4 Problem Formulation . . . 34
3.2.5 Neural Network Training . . . 36
3.2.6 Benchmark Algorithms . . . 37
3.3 Neural Networks for Precoding . . . 38
3.3.1 Suitable Neural Network Architectures . . . 38
3.3.2 Graph Neural Network Precoding . . . 42
3.3.3 Leveraging SNR Regime Information . . . 44
3.4 Complexity and Real-Time Analysis . . . 45
3.4.1 Complexity of Zero Forcing Precoder . . . 47
3.4.2 Complexity of Distortion-aware Beamforming Precoder . . . 48
3.4.3 Numerical Comparison . . . 49
3.5 Simulation Results . . . 50
3.5.1 Single-User Case . . . 51
3.5.2 Multi-User Case . . . 54
3.5.3 Improvements in Power Consumption . . . 57
3.5.4 Validation on Rapp PA Model . . . 61
3.5.5 Leveraging SNR Regime Information . . . 61
3.6 Conclusion . . . 63

4 Generalization to Varying Channel Conditions 64

4.1 Introduction . . . 64
4.2 Problem Formulation . . . 65
4.2.1 System Model . . . 66
4.2.2 GNN-based Precoder . . . 66
4.2.3 Channel models . . . 67
4.3 Simulation Results . . . 68
4.3.1 Methodology . . . 68
4.3.2 Rayleigh Training Distribution . . . 68
4.3.3 Mixed Training Distributions . . . 70
4.4 Conclusion . . . 72

5 Extensions to Multi-Carrier Systems 74
5.1 Introduction . . . 75
5.2 Problem Formulation . . . 76
5.2.1 System Model . . . 76
5.2.2 Channel Models . . . 77
5.2.3 Power Amplifier Model . . . 78
5.2.4 Achievable Rate . . . 79
5.3 GNN Based-precoding . . . 80
5.3.1 Frequency-flat Neural Network Precoder . . . 81
5.3.2 Frequency-selective Neural Network Precoder . . . 82
5.3.3 Loss Functions . . . 82
5.4 Inference Complexity . . . 83
5.4.1 Classic Precoding . . . 83
5.4.2 GNN Frequency-flat Channel . . . 83
5.4.3 GNN Frequency-selective Channel . . . 84
5.5 Power Consumption . . . 85
5.5.1 PA Consumption . . . 85
5.5.2 GNN Inference Consumption . . . 87
5.6 Simulation-based Performance Assessment . . . 88
5.6.1 Results Fixed IBO . . . 88
5.6.2 Results Varying IBO . . . 90
5.6.3 Power Consumption Comparison . . . 95
5.7 Out-of-band Emissions and Oversampling . . . 99
5.7.1 Oversampling . . . 100
5.7.2 In-band Performance with Oversampling . . . 102
5.7.3 Out-of-band Performance . . . 105
5.8 Conclusion . . . 109

6 Learning to Quantize & Precode 111
6.1 Introduction . . . 113

6.1.1 Problem Statement . . . 113
6.1.2 State-of-the-Art Quantization in Massive MIMO . . . 114
6.1.3 State-of-the-Art Neural Network-based Precoding . . . 115
6.2 System Model . . . 117
6.2.1 Linear Quantized Precoding . . . 117
6.2.2 Quantization . . . 118
6.2.3 Non-Linear Neural Network-based Quantized Precoding . . . 120
6.2.4 Achievable Sum Rate . . . 120
6.2.5 Radiation Patterns . . . 121
6.3 Spatial Characteristics of Quantization-Induced Distortion . . . 122
6.4 Neural Networks for Precoding . . . 122
6.4.1 General Training Procedure . . . 123
6.4.2 Unsupervised Learning of Non-Linear Precoding . . . 125
6.4.3 Neural Network Architecture . . . 129
6.5 Complexity Analysis . . . 131
6.5.1 Complexity of the GNN . . . 131
6.6 Simulation-based Performance Assessment . . . 133
6.6.1 Training Parameters and Simulation Setup . . . 133
6.6.2 GNN-Based Non-Linear Precoding Single-User Case . . . 134
6.6.3 GNN-Based Non-Linear Precoding Multi-User Case . . . 138
6.7 Power Consumption . . . 138
6.7.1 DACs Power Consumption . . . 139
6.7.2 GNN Processing Power Consumption . . . 140
6.8 Practical Implementation Challenges . . . 143
6.9 Conclusion . . . 143

7 Valorization **145**
7.1 Economic and Ecological Alignment . . . 145
7.2 AI Potential and Limitations in the Physical Layer . . . 147

8 Conclusion and Future Work **153**
8.1 Conclusion . . . 153
8.2 Future Work . . . 156

Bibliography **159**

List of Publications **175**

List of Figures

1.1 Simplified overview of a massive MIMO system. 3
1.2 Massive MIMO system with hardware impairments. 4
1.3 Overview of thesis chapters. 10

2.1 Typical power amplifier characteristic. 15
2.2 Heatmap of the distortion at all observer locations. 20
2.3 Third-order distortion pattern of maximum ratio transmission (MRT) vs. zero third-order distortion (Z3RO). 22
2.4 Rate in function of signal-to-noise ratio (SNR) for MRT and Z3RO. 22
2.5 Empirical cumulative distribution function (eCDF) of the distortion level for the single and multi-user case. 23

3.1 Massive MIMO downlink with neural network-based linear precoding and non-linear amplifiers at the transmitter. 30
3.2 Amplitude modulation to amplitude modulation (AM/AM) and amplitude modulation to phase modulation (AM/PM) characteristics of the considered power amplifier models. 33
3.3 Hypothesis spaces of different neural network architectures, assuming infinite-capacity models. 41
3.4 Comparison of performance and number of trainable parameters for multilayer perceptron (MLP), circular convolutional neural network (CCNN) and graph neural network (GNN) precoding. 42
3.5 GNN visualized for a single edge in the graph. 45
3.6 Achievable rates evaluated on a 3^{rd} order power amplifier (PA) (a) and an 11^{th} order PA (b). 52
3.7 Amplitude and phase plot per antenna, GNN compared to the Z3RO precoder. 53
3.8 Radiation pattern of the intended signal, distortion signal and signal-to-distortion ratio (SDR) for $K = 1$ user. 54

3.9 Achievable rates evaluated on an 11^{th} order PA. Comparing the GNN precoder against zero forcing (ZF) and the distortion-aware beamforming (DAB) precoder. 55
3.10 Radiation pattern of the intended signal, distortion signal and SDR for $K = 2$ users. 57
3.11 Radiation pattern of the intended signal, distortion signal and SDR for $K = 4$ users. 58
3.12 (a) Consumed PA power in function of input back-off (IBO) (b) achievable sum rate in function of consumed PA power. 59
3.13 Achievable sum rates evaluated using the Rapp PA model. 62
3.14 Achievable rates for $M = 64$, $K \in \{1, 2, 4\}$ and an 11^{th}order PA. Comparing GNN base against the GNN that leverages SNR information. 62

4.1 Magnitude of the different channel distributions. 68
4.2 Achievable rate for $\text{GNN}_{\text{Rayleigh}}$ and GNN_{LoS}. 69
4.3 Achievable rate for $\text{GNN}_{\text{Rayleigh}}$ and GNN_{Rice} 70
4.4 Achievable rate for $\text{GNN}_{\text{Rayleigh}}$ on a line-of-sight (LoS) and Rician testset. 71
4.5 Difference in achievable sum rate of $\text{GNN}_{\text{Rician}}$ and GNN_{LoS} compared to $\text{GNN}_{\text{Rayleigh}}$ 71
4.6 The performance degradation is expressed as $\Delta R_{\text{sum}} = R_{\text{sum,benchmark}} - R_{\text{sum,GNN}_{\text{Rayleigh}}}$ In (a) the considered benchmark is ZF, tested on Rician and LoS channels. In (b) the considered benchmark is $\text{GNN}_{\text{mixed}}$ tested on Rayleigh, Rician and LoS channels. The considered test set is indicated between brackets in the legend. . . 72

5.1 Graph representing the wireless system. 80
5.2 Rate vs SNR_q on strongly frequency selective/independently and identically distributed (i.i.d.) and frequency flat channels. 89
5.3 Rate vs β on a channel with varying degrees of frequency selectivity. 90
5.4 Comparing the average rate and PA consumed power of the different precoders in function of β for a number of IBO values. . 94
5.5 Herfindahl–Hirschman index (HHI)-score of the different precoders at different IBOs as a function of the channel frequency-selectivity (β). 95
5.6 Amplitude and phase plot per antenna for a single subcarrier, comparing GNN_Q with MRT. 96
5.7 Required $p_{\text{cons,PA}}$ (a) and IBO_{max} (b) to meet $R \geq 4$ in function of β. 98

5.8 Required $p_{\text{cons,PA}}$ (a) and IBO_{max} (b) to meet $R \geq 8$ in function of β. 99
5.9 Required $p_{\text{cons,PA}}$ (a) and IBO_{max} (b) to meet $R \geq 11$ in function of β. 99
5.10 Rate vs SNR_q on frequency flat channels with oversampling. . . . 103
5.11 Rate vs SNR_q on strongly frequency selective/i.i.d. channels with oversampling. 104
5.12 Effect of oversampling on rate vs β. 105
5.13 Best- and worst-case power spectral density (PSD) across the antennas for MRT and GNN precoding for frequency-flat $\beta = 100$ (a) and frequency selective channels $\beta = 10^{-3}$ (b). With $Q = 128$, average $\text{IBO} = -3$ dB and $L_{\text{os}} = 5$. 107
5.14 Spatial adjacent channel leakage ratio (ACLR), $\text{ACLR}(\theta')$ [dB] for MRT and GNN precoding for $Q = 128$, average $\text{IBO} = -3$ dB and $L_{\text{os}} = 5$. Evaluated on a LoS channel, where the user is indicated by a yellow star. 108

6.1 Simplified system overview of linear quantized precoding. 117
6.2 Simplified system overview of non-linear quantized precoding. . . . 117
6.3 Radiation pattern of the intended signal, distortion signal and SDR for a 1-bit DAC and MRT precoding for different user angles. . . . 123
6.4 Radiation pattern of the intended signal, distortion signal and SDR for a 1-bit DAC and ZF precoding for different user angles. 123
6.5 Overview of the forward and backward pass during training. 124
6.6 GNN edge update overview. 128
6.7 Achievable rates of the GNN non-linear precoder, MRT or ZF, for the single-user case and a varying number of bits (a) and the multi-user case with one bit (b). 133
6.8 Radiation pattern of the intended and distortion signal for a 1-bit DAC. The GNN precoder (a) and MRT precoding (b) are compared. In (c) the signal-to-distortion ratio (SDR) radiation pattern is depicted. 136
6.9 Radiation pattern of the GNN for the intended and distortion signal for a 1-bit DAC, varying number of transmit antennas and $K = 1$ user. 137
6.10 Real part of the estimated symbol in function of the transited symbol, over a single realization of a noiseless Rayleigh fading channel. Comparing MRT with GNN precoding. 137
6.11 Achievable rates, comparing GNN and MRT or ZF for $M = 32$, $K \in \{1, 2, 4, 6\}$. 139

6.12 Power consumption of the digital-to-analog converters (DACs) and GNN processing in function of the bandwidth. 142
6.13 Power consumption of the radio frequency digital-to-analog converters (RF-DACs) and GNN processing in function of the bandwidth. 142

List of Tables

1.1 Energy consumption breakdown of 4G [1] and 5G [2] base stations. 6

2.1 Overview of the open-source measurements. 19

3.1 Complexity of the GNN, ZF and DAB precoders. 49
3.2 PA parameters of the 11^{th} order polynomial PA model at different back-off values ($\beta_1 = 1$). 50
3.3 Consumed power of the GNN, ZF and ZF + DPD precoders for $M = 64$, $K = 4$, for a fixed data rate of 30 bits/channel use. . . . 61

5.1 Overview of the different precoders, including feature sizes, complexity, required speed and inference power consumption. . . . 84

6.1 Normalized mean squared error (NMSE) for MRT and GNN on all (noiseless) channels and symbols from the test set, for $M = 32$, $K = 1$. 136

Chapter 1

Introduction

This thesis investigates how energy reduction can be achieved in massive multiple input multiple output (MIMO) base stations (BSs) by efficiently coping with hardware nonlinearities in the transceiver chain. The focus is on two key contributors to the power consumption, namely the power amplifiers (PAs) and the digital-to-analog converters (DACs), whose operation is traditionally constrained by linearity requirements. By relaxing these constraints and tolerating controlled nonlinear distortion, this thesis explores learning-based precoding strategies that enable operation closer to the hardware limits while maintaining high data rates and acceptable signal quality. Section 1.1 provides a general motivation for this work, while Section 1.2 gives an introduction to massive MIMO systems, the energy consumption of wireless networks and the role machine learning (ML) can play in these systems. Section 1.3 provides an overview of the research questions. Finally, Section 1.4 outlines the remainder of this thesis.

1.1 From Maximum Performance to Power-Aware Design

Over the last few decades, the demand for high-throughput wireless communications has grown tremendously. More specifically, in the last decade, global network traffic has more than doubled every two years, resulting in a global data traffic increase of 300 times between 2011 and 2021 [3]. This growth has slightly slowed down, however, even today, data traffic shows a year-to-year growth of 20% [4]. Unfortunately, this ever-growing demand for high data rate communications comes

with an increase in energy consumption [5]. Given the past annual data traffic growth of 50%, the energy efficiency of wireless networks should improve by nearly 60 times every 10 years, in order to maintain a stable energy consumption [6]. While energy efficiency has increased over the different mobile network generations, it has generally been a low-priority factor as compared to classical performance indicators, such as high data rates and capacity [7]. The focus on improving performance is driven by this drastic increase in data traffic demand. Consequently, this has resulted in an increase in energy consumption of mobile networks [8, 5]. This growing energy consumption contributes to the growing carbon footprint of the ICT sector [9]. This is a problem given the strict ambitions on sustainability that have been formulated, such as the ones put forward in Europe's Green Deal [10].

A clear example of prioritizing performance (i.e., high signal quality and thus data rate) over power-aware design can be seen in the PAs in wireless systems, as they are one of the major energy consumers of BSs [11]. These PAs are typically operated in a linear regime where they do not distort the amplified signal. However, this comes at the cost of a reduced energy efficiency of the PAs. Consequently, PAs are operated at efficiencies as low as $\sim$ 5-30% [1, 12]. Another such example comes from analog-to-digital converters (ADCs) and DACs. It is known that their power consumption scales exponentially with the number of bits used and linearly with the considered bandwidth [13, 14]. Nevertheless, current DACs and ADCs are typically designed for a high number of bits, raising their power consumption. On top of this, wireless communication systems are moving towards higher carrier frequencies and, as such, higher bandwidths, resulting in a further increase in power consumption [15, 13].

In recent years, a paradigm shift has been gaining momentum, where instead of designing for maximum performance, a trade-off with minimizing power consumption is considered in the design of wireless systems [7, 16, 17]. This paradigm shift is driven by growing environmental concerns and increased energy costs. Notably, taking energy efficiency into account in the optimization of wireless systems often leads to complex problems. This complexity arises from operating the hardware in energy-efficient conditions, where unfavorable characteristics occur, such as nonlinearity, non-differentiability and non-continuity. These complex problems are not easily solved using classical techniques. As such, this provides fertile ground for research. Artificial intelligence (AI) and ML are tools that show great potential to tackle such complex problems. These promising techniques, however, come with an important caveat: in modern machine learning, improved performance is frequently achieved through increased model size. As such, their computational complexity, and strongly linked to this, their energy consumption can be substantial. Consequently, the use of such tools comes with an additional

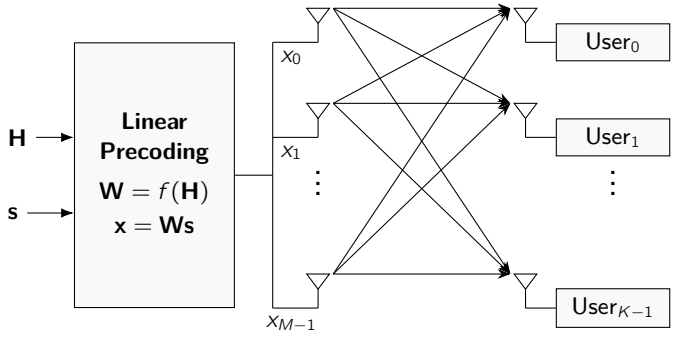


Figure 1.1: Simplified overview of a massive MIMO system.

power consumption that has to be accounted for.

1.2 Terminology and Preliminaries

1.2.1 Massive MIMO: a Blessing or a Curse?

The tremendously growing demand for high-throughput wireless communication has led to the development of the 5G standard, which is currently powered by many technologies. One of the main technologies is massive MIMO, which many telecom operators have started to utilize to support 5G [18]. Massive MIMO uses many antennas at the BS combined with clever precoding to direct the signal towards the intended users. Precoding changes the phase and amplitude of the signals transmitted by each antenna to spatially direct the signal towards the users (rather than sending an omnidirectional signal when using a single antenna) and to avoid user interference. Generally speaking, precoding results in constructive combining at the user location (i.e., the signals from the different antennas add up). By focusing the signals in the user direction, several users can be served at the same time by separating them in space, increasing the network capacity and spectral efficiency [19].

We assume a massive MIMO base station equipped with M antennas that serves K single-antenna users simultaneously over the same time and frequency resources. In this work, the downlink scenario is considered where transmission is from the BS to the users. This system is illustrated in Figure 1.1. Theoretically speaking, the use of this system has many benefits, such as (*i*) array gain, (*ii*) channel hardening, and (*iii*) favorable propagation. (*i*) In a massive MIMO system, precoding enables

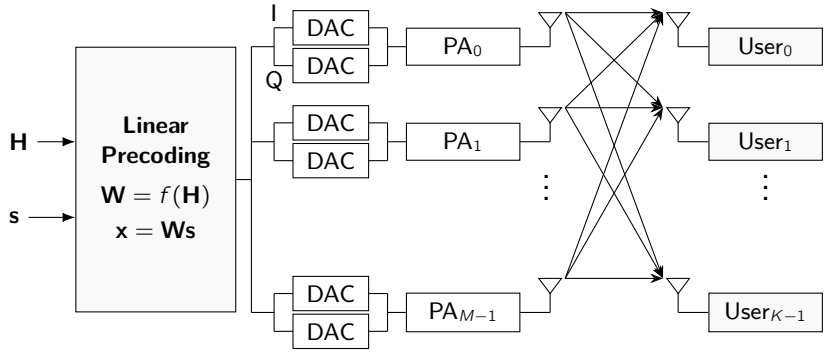


Figure 1.2: Massive MIMO system with hardware impairments.

the signals from the many transmit antennas to coherently combine at the user location [20]. This constructive interference of the transmit signals results in an increased signal-to-noise ratio (SNR) at the user location. This increased SNR scales with the number of transmit antennas M and is referred to as the array gain. (*ii*) A typical wireless channel can have large fluctuations in channel gain over time and frequency due to multi-path propagation, which is called small-scale fading. By utilizing many transmit antennas and adequate precoding, these fluctuations average out, resulting in the instantaneous channel gain across the array behaving nearly deterministic. More specifically, due to the law of large numbers, as the number of transmit antennas M increases, the instantaneous channel gain across the array converges to its deterministic average channel gain [20]. This phenomenon is known as channel hardening. (*iii*) When the BS communicates to multiple users at the same time and frequency, this will create user interference. To minimize interference the user channels should be sufficiently different. In massive MIMO, favorable propagation refers to the regime where the channel vectors of different users become asymptotically orthogonal as the number of BS antennas grows. Under this condition, inter-user interference naturally diminishes, spatial multiplexing becomes highly efficient, and simple linear processing achieves close to optimal performance [19].

These benefits are all driven by the scaling of the number of transmit antennas/chains. However, these benefits come with significant challenges. The hardware complexity associated with hundreds of radio frequency chains and PAs introduces new energy and cost constraints. Nonlinear distortions, PA and DAC inefficiencies, and RF impairments can collectively degrade performance and offset the theoretical gains. This may lead to a higher total energy consumption due to more circuitry, inefficient hardware, and complex signal processing. A

more complete system overview of a massive MIMO system that contains the impairments considered in this work is illustrated in Figure 1.2.

In summary, while massive MIMO enables a reduction in transmit power for the same quality of service (QoS), due to the array gain, the added consumption of the additional radio frequency chains and complex processing may outweigh these gains. Consequently, massive MIMO enables unprecedented levels of network performance but simultaneously demands novel approaches in system design, signal processing, and hardware implementation.

1.2.2 Energy Consumption of Wireless Networks

In this section, starting at a high level, the energy consumption of the information and communication technologies (ICT) sector is outlined. Next, we go further into communication networks, mobile networks, base stations and the respective components of the base stations. This leads us to the PA and DAC energy consumption, which are the focus areas of this work.

ICT Sector: The ICT sector is typically divided into three parts: networks, data centers, and user devices. Looking at the ICT sector as a whole, its energy consumption is roughly 4-7 % of the global electricity consumption [21, 9, 22, 5, 23]. This contributes to approximately 2-4 % of the global carbon footprint [21, 9, 22, 5, 23]. To put this into perspective, the ICT sector's carbon emission is similar to the emissions of the global aviation industry, which accounts for approximately 2-2.5 % of the global carbon emissions [24, 25]. In the next paragraph, we zoom in on the network component.

Communication Networks: Within the ICT sector, communication networks represent a substantial amount of the emissions. Estimates vary, but there is a consensus that 20-35 % of the ICT emissions stem from communication networks [9, 22, 5]. Zooming in further, approximately 60 % of the electricity consumption of communication networks is accounted for by mobile networks [23]. Mobile networks are constructed of the core network and the radio access network. The radio access network consists of a large number of BSs that provide wireless access to user devices. Typically, 80 % of the energy consumption of mobile networks stems from the radio access network/base stations [2].

Base Stations: Given the large share of energy consumption/carbon emissions the BSs account for within wireless networks, they are the main focus of this work. The energy consumption of BSs is typically divided into four major components: the PAs, the analog front-end (AFE), the digital baseband (DBB) and the power supply and cooling systems (PSC). The AFE consists of all components in the radio frequency (RF) chain that perform analog operations, such as the ADC, DAC, up and down-conversion, clock generation, etc. [2]. Conversely, the DBB performs all digital functions such as modulation, precoding, decoding, demodulation, etc. [2].

Table 1.1: Energy consumption breakdown of 4G [1] and 5G [2] base stations.

Generation	PAs	AFE	DBB	PSC
4G	20-60%	5-12%	10-50%	15-30%
5G	35-55%	20-40%	10-15%	10-15%

In 4G systems, the energy consumption is distributed over these components as illustrated in Table 1.1. These numbers vary depending on the size of the BS and the load. From this, it is clear that the PAs are the largest energy consumers within 4G BSs. In more recent 5G systems, a shift has occurred; the energy consumption breakdown is illustrated in Table 1.1. Notably, the PAs are still one of the major consumers, however, the AFE has an increased energy consumption. This is due to the increase in RF chains considered in 5G systems. As the PA is still the major consumer within BSs, reducing its energy consumption is highly valuable. Additionally, the energy efficiency of the PA is typically low, due to the energy-efficiency/linearity trade-off [12], making it a suitable target for optimization. Next to this, the aim of this work is to reduce the power consumption of the DAC, which is part of the AFE, which incurs an increased energy consumption in 5G systems. The use of many radio frequency chains increases both the hardware complexity and the power consumption. Additionally, the current move towards higher carrier frequencies and higher bandwidths [15], further increases the complexity and the power consumption of the DACs. This is because the power consumption of typical current steering DACs scales exponentially with the number of bits and linearly with the sampling frequency [13].

1.2.3 Opportunities of Artificial Intelligence and Machine Learning

Motivated by the high energy consumption of wireless networks, the aim of this work is to optimize certain components in the BS to reduce their energy consumption. This, however, often leads to intractable, complex, non-convex optimization problems, which are challenging to solve using classical signal processing techniques. For instance, operating the PA more efficiently corresponds to introducing nonlinear distortion into the system. This is challenging for classical signal processing techniques, which commonly rely on linearity assumptions. Another example is the DAC, using fewer bits reduces its energy consumption. However, this increases the quantization-based distortion, which traditionally could be neglected as sufficient bits would be used. Taking the quantization into account is challenging as it is a nonlinear, non-differentiable, and non-continuous function.

In sum, optimizing for energy efficiency oftentimes results in breaking classical assumptions in order to consider more realistic hardware models. This leads to challenging problems, where it is non-trivial to apply classical (linear) signal processing techniques. Even if traditional techniques can be applied, the non-convexity of the problems frequently leads to a significant computational complexity cost.

These challenging problems open the door for ML and AI methods for two reasons. First, ML can tackle highly non-convex optimization problems as learning the solution structure from data is generally easier than hand-designing a suitable algorithm. Second, a lot of the computational complexity is offloaded to the training step, which reduces inference complexity. Coupled with the highly parallel structures of neural networks (NNs) this can lead to low-complexity solutions which can benefit from a fast parallel implementation. Although a parallel implementation can lead to fast inference times, we need to be wary of the associated energy cost of the introduced computations, such that they do not outweigh the energy reduction achieved in the PAs/DACs.

1.3 Research Questions

The overall goal of this PhD thesis is to reduce the energy consumption of massive MIMO systems, while maintaining the QoS. This goal is addressed by developing solutions that enable efficient operation of hardware components in the transmitter that consume a lot of power. More specifically, the aim is to reduce the power

consumption of PAs and DACs. Operating these hardware components in an energy-efficient manner introduces certain nonlinear distortions that have to be managed. This leads to complex non-convex optimization problems. As such, the use of NNs is studied to tackle these problems. The research questions are split up into two main groups. The first group is related to the introduced nonlinearities and how to overcome these. These research questions are formulated as follows:

Q1 - Which classical precoding techniques exist to mitigate nonlinear PA distortion and what are their limitations?

Q2 - Can NN-based precoding be leveraged to allow PAs to operate close to saturation, while maintaining the data rate?

- Q2.1 - Which neural network architectures are suitable for linear precoding?
- Q2.2 - What is the impact of varying channel conditions (different training distribution vs. testing distribution) on the performance of NN-based precoders?
- Q2.3 - In a multi-carrier system, where inter-carrier interference occurs due to the nonlinear PA functions, how can NN-based distortion-aware precoding be used/adapted to allow PAs to operate close to saturation while maintaining a high data rate?

Q3 - How can NN-based nonlinear precoding architectures be designed such that the number of DAC bits can be reduced without degrading the data rate?

Questions 1 and 2 are related to the power amplifier. Question 1 is addressed in Chapter 2 and motivates the need for NN-based precoding techniques. These techniques address Questions 2 and its sub-questions and are developed in Chapters 3, 4 and 5. Question 3 explores the option to use a similar methodology, i.e., a NN-based precoder, to tackle quantization-induced distortion in DACs and is covered in Chapter 6.

The second group of research questions relates to the potential reduction in energy consumption that can be achieved by operating under the nonlinear but more energy efficient conditions, i.e., operating the PAs close to saturation and operating the DACs with fewer bits. These research questions are outlined as follows:

Q4 - What is the potential power consumption reduction of the developped NN-based precoders?

Q4.1 - How much can the PA consumed power be reduced, and does this outweigh the additional inference power consumption of the required NN?

Q4.2 - How much can the DAC consumed power be reduced, and does this outweigh the additional inference power consumption of the required NN?

Question 4.1 considers the PA consumed power and assesses the global power reduction, taking into account the reduction in PA consumed power and the increase in digital signal processing consumed power, introduced by the inference power consumption of the used NN. This question is addressed in Chapters 3 and 5. Question 4.2 relates to the reduced power consumption of the DAC, which is achieved by using fewer bits. Again, the global power consumption is considered, which accounts for the reduction in consumed power of the DACs and the increased digital signal processing consumed power, introduced by the inference power consumption of the used NN. This question is addressed in Chapter 6.

1.4 Thesis Outline and Contributions

This section provides an overview of the remainder of this thesis and makes the link with the contributions of this research, which have been published in different journal and conference papers. An overview of the different chapters, the link with the research questions and the contributions can be seen in Figure 1.3. Next, a concise summary of the contributions and key findings of each of the following chapters is provided.

Chapter 2 - Power Amplifier Nonlinearities & Analytical Solutions. This chapter provides an introduction to the problem of nonlinear PA distortion. It outlines classical approaches to tackle the problem. In particular, it examines the performance of the Z3RO precoder, which is a closed-form distortion-aware precoder. The performance gains under measured channels are illustrated. The limitations of the Z3RO precoder are denoted, forming the motivation for the following chapters.

Chapter 3 - Neural Network Approaches for Distortion-Aware Precoding. Chapter 3 proposes the use of NNs for distortion-aware precoding. By utilizing NNs more realistic conditions can be considered, such as higher-order polynomial

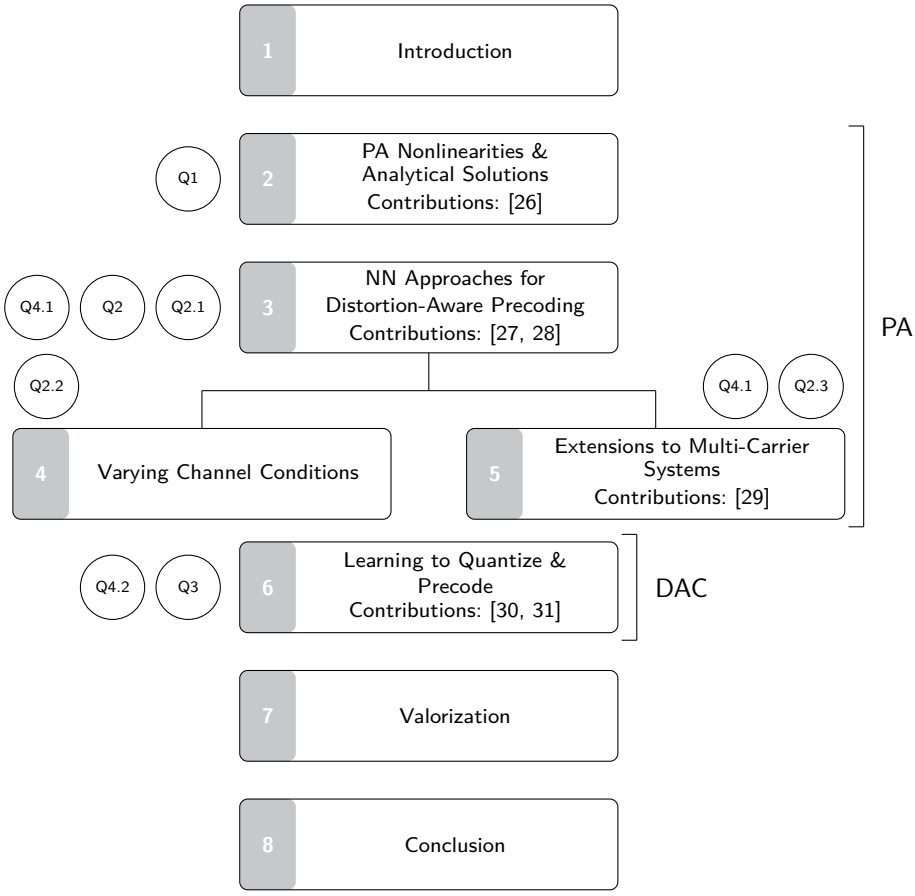


Figure 1.3: Overview of the thesis chapters, linked with the research questions and contributions. The chapters related to the PA and DAC are indicated. Some chapter titles are shortened for readability.

PA models and multi-user scenarios. It is shown that this NN-based precoder can achieve a high data rate under non-linear distortion, enabling energy-efficient operation of the PAs. Additionally, the overall energy reduction is computed, considering both the energy reduction in the PAs and the energy increase in the digital signal processing required for the NN.

Chapter 4 - Generalization to Varying Channel Conditions. The NN precoder in Chapter 3 is trained assuming a channel model. When there is a difference between the training distribution and the testing distribution, a performance reduction can be expected; this phenomenon is known as domain or covariate shift. In this chapter, it is explored whether covariate shift results in a performance reduction, and if so, how severe the performance drop is.

Chapter 5 - Extensions to Multi-Carrier Systems. The NN-based precoder developed in Chapter 3 considers narrow band channels, i.e., frequency-flat channels. To extend the proposed solution to a frequency-selective scenario, a multi-carrier system is considered. This extension is non-trivial due to the presence of the nonlinear PA functions, as they cause inter-carrier interference. In this chapter, this problem is tackled by proposing a NN-based precoder that jointly precodes across all subcarriers. Additionally, the overall energy reduction is computed, considering both the energy reduction in the PAs and the energy increase in the digital signal processing required for the NN.

Chapter 6 - Learning to Quantize & Precode. In chapter 6, the problem of quantization-induced distortion is considered. This problem arises when DACs are selected with few bits. Using fewer bits allows for a reduced energy consumption of the DACs. In order to handle the quantization distortion a nonlinear NN-based precoder is proposed. To assess the overall energy reduction both the energy reduction in the DACs and the energy increase in the digital signal processing required for the NN are considered.

Chapter 7 - Valorization. This chapter covers the valorization potential of the performed research. It discusses how learning-based precoding for massive MIMO systems with hardware nonlinearities can contribute to energy savings in wireless networks, and how these savings align with economic and societal objectives. Next to this, the opportunities and limitations of AI/ML for physical layer communications are discussed, highlighting the next steps for future deployability.

Chapter 8 - Conclusion. In this final chapter, an overview of the contributions, along with an overall conclusion, is provided. The conclusion formulates an answer to the proposed research questions. Next to this, future directions are discussed that could strengthen and extend the results of this thesis.

Chapter 2

Power Amplifier Nonlinearities & Analytical Solutions

This chapter is based on the following paper:

> T. Feys, G. Callebaut, L. Van der Perre, and F. Rottenberg, "Measurement-Based Validation of Z3RO Precoder to Prevent Nonlinear Amplifier Distortion in Massive MIMO Systems," in *2022 IEEE 95th Vehicular Technology Conference: (VTC2022-Spring)*, pp. 1–5, 2022 [26]

In this chapter, the problem of power amplifier (PA) nonlinearity is introduced. This is tightly coupled to the linearity/energy-efficiency trade-off of the PA. Classical solutions to PA nonlinearities are outlined and their limitations are discussed. In particular, the potential gain of the Z3RO precoder is examined and validated based on real-world channel measurement data. The Z3RO precoder is an analytical distortion-aware precoder aimed at nulling the PA distortion in the user direction, which is developed for the single-user case. This chapter aligns with research question (Q1) 'Which classical precoding techniques exist to mitigate nonlinear PA distortion and what are their limitations?'. It is shown that the Z3RO precoder can adequately reduce the distortion experienced by the user by sending it in non-user directions. However, the Z3RO precoder relies on simplifying assumptions in order to provide an analytical solution. For instance, it assumes a third-order polynomial PA model, which is only valid as the PA enters saturation. To model deep

saturation, higher-order polynomial models for the PA are required. Additionally, the Z3RO precoder is designed to serve a single user. Extensions to more realistic PA models and multi-user support form the motivation for the work in the following chapters.

The remainder of this chapter is outlined as follows. In Section 2.1, the problem of PA nonlinearity and the trade-off with energy efficiency are introduced. Additionally, conventional solutions and their limitations are outlined. Next, Section 2.2 details the system model used to perform the simulations. Section 2.3 discusses the simulation results. Finally, Section 2.4 concludes the chapter.

2.1 Introduction

2.1.1 Problem Statement

In current wireless cellular communication systems, the PA is one of the main sources of energy consumption [1, 2]. To achieve good communication performance, i.e., a high quality of service (QoS), the PAs should be highly linear. However, there is a trade-off between power-efficiency and linearity of the PA [32]. This trade-off is illustrated in Figure 2.1, where a typical PA characteristic is depicted. For small input and output powers, the PA operates in the linear regime, where little nonlinear distortion is experienced. However, this comes at the cost of a low energy efficiency of the PA. As the input power rises, the PA enters saturation. This results in a higher energy efficiency, but unfortunately introduces nonlinear distortion due to the compression of the output signal. Consequently, when using a PA at an energy-efficient operating point, nonlinear distortion will arise. This nonlinear distortion is detrimental for the performance of the wireless system and can furthermore lead to unauthorized out-of-band (OOB) emissions. This is especially the case in massive multiple input multiple output (MIMO) systems, where beamforming allows the base station to transmit the signal in a certain spatial direction. Studies have shown that the nonlinear distortion is not spatially spread out, but follows the dominant beamforming direction [33, 34]. This is witnessed in particular in situations with predominant transmission in one or a few directions (e.g. line-of-sight (LoS) situations with only few users). This beamforming of the nonlinear distortion to the user location can strongly decrease the signal-to-distortion ratio (SDR) at the user location, which inherently limits the user performance. Additionally, this can introduce or strengthen unauthorized OOB emissions at both the user and unintended locations.

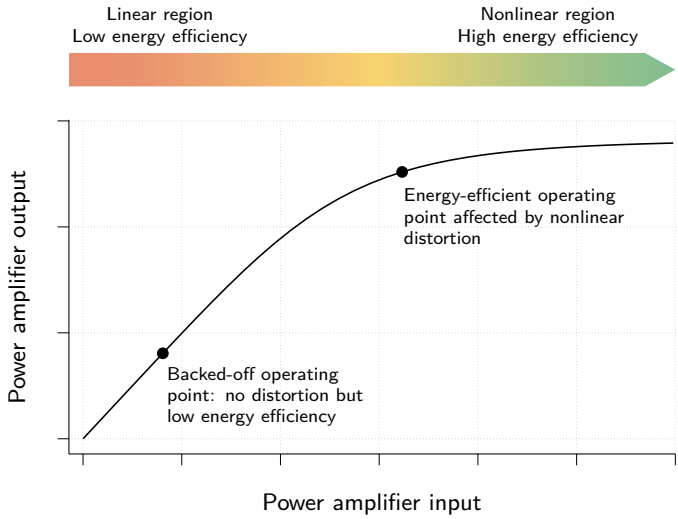


Figure 2.1: Typical PA characteristic showing the nonlinearity and efficiency of the PA [12].

2.1.2 Conventional Solutions and their Limitations

One approach to avoid nonlinear distortion is to use the PA with a certain back-off power, i.e., reduce the transmit power to set the operating point of the PA sufficiently far away from the saturation point. This can be seen in the left operating point in Figure 2.1. This method avoids the PA operating in the nonlinear region, hence avoiding nonlinear distortion. However, increasing the back-off reduces the energy-efficiency of the PA. Another drawback is the fact that reducing the transmit power can lead to poor signal-to-noise ratio (SNR) when the user experiences a high path loss.

An alternative solution to the problem is to use digital pre-distortion (DPD) techniques [32]. These techniques pre-distort the unamplified signal to compensate for the nonlinear behavior of the PA. This pre-distortion results in the amplified signal being a linear amplification of the original signal. Unfortunately, DPD techniques introduce an additional complexity burden, which is especially present in massive MIMO systems where the DPD has to operate on a large number of antenna chains. Another limitation of DPD is the fact that it can only account for weak nonlinear effects. Even a perfect DPD can only linearize the PA function up to the saturation point of the PA. This is especially problematic for high peak-to-average power ratio (PAPR) signals such as orthogonal frequency division

multiplexing (OFDM) signals [35], which are typically used in current 5G massive MIMO systems.

More recent works have looked at how to reduce this nonlinear distortion by taking it into account in the precoding design [36, 37, 38]. These works rely on iterative procedures to allow the PA to work closer to saturation and/or to alleviate the need for DPD. In [39], the Z3RO precoder was derived which circumvents the need for an iterative approach as a closed-form solution has been derived which maximizes the SNR while nulling the third-order distortion term.

2.1.3 Contributions

In [39], the authors proposed a novel precoder that allows the PA to work close to saturation, while canceling the third order nonlinear distortion at the user location. The zero third-order distortion (Z3RO) precoder is obtained by maximizing the SNR, while constraining the third-order distortion at the user location to zero. The precoder is obtained by saturating a subset of the antennas and inverting the sign of their phase shift. In this chapter, the Z3RO precoder (originally designed for the single-user case) is evaluated in the multi-user case. Additionally, given the novelty of the Z3RO precoder, measurement-based simulations are performed to further evaluate its performance. The open-source channel measurements presented in [40] are used as the basis for a simulation-based assessment. It is possible to calculate the distortion value at all measured locations, using these channel measurements. The performance of the Z3RO precoder is assessed based on these simulations. A comparison between the Z3RO precoder and maximum ratio transmission (MRT) is made, given that MRT is the optimal precoder for the single-user case when no distortion is present. Additionally, for the multi-user case only a small amount of users is considered given that no interference-cancellation is present in the Z3RO and MRT precoders.

2.2 Transmission Model

2.2.1 Transmit Signal

We consider a massive MIMO system where the base station (BS) is equipped with M antennas and K single-antenna users are served using the same time and frequency resources. The complex transmit signal for user k, denoted as s_k, is assumed to be complex Gaussian distributed with variance p_k, and uncorrelated

with the signal of other users. The signal is precoded at antenna m, using the coefficient $w_{m,k}$. The signal before the PA is thus represented as

$$x_m = \sum_{k=0}^{K-1} w_{m,k} s_k.$$

For the sake of clarity and without loss of generality, the PA gain is set to one.

2.2.2 Z3RO and MRT Precoders

For MRT precoding, the complex precoding weights are matched to the channel coefficients in order to maximize the SNR at the intended users. The Z3RO precoder uses a number of saturated antennas M_s to compensate for the distortion caused by all other antennas, improving the SDR at the intended users.

The total power at the input of all the PAs is $p_T = \mathbb{E}\left(\sum_{m=0}^{M-1} |x_m|^2\right) = \sum_{k=0}^{K-1} p_k \sum_{m=0}^{M-1} |w_{m,k}|^2$. The average power at the input of each PA is $p_{\text{in}} = p_T/M$, with $p_{\text{in}} = \sum_{k=0}^{K-1} p_k$. In order to respect this power budget, a normalization factor α_k ensures that $\sum_{m=0}^{M-1} |w_{m,k}|^2 = M$. Hence, the average back-off at each PA becomes $p_{\text{in}}/p_{\text{sat}}$, where p_{sat} is the saturation power of the PA.

The classical MRT precoder is defined as

$$w_{m,k}^{\text{MRT}} = \alpha_k^{\text{MRT}} h_{m,k}^*,$$

where

$$\alpha_k^{\text{MRT}} = \sqrt{\frac{M}{\sum_{m'=0}^{M-1} |h_{m',k}|^2}}$$

normalizes the power of each user to M.

The Z3RO precoder [39] is defined as

$$w_{m,k}^{\text{Z3RO},M_s} = \alpha_k^{\text{Z3RO}} h_{m,k}^* \begin{cases} -\gamma_k & \text{if } m = 0, \ldots, M_s - 1 \\ 1 & \text{otherwise} \end{cases},$$

where M_s is the number of saturated antennas with an opposite phase, γ_k is the additional gain of these saturated antennas given by

$$\gamma_k = \left(\frac{\sum_{m'=M_s}^{M-1} |h_{m',k}|^4}{\sum_{m''=0}^{M_s-1} |h_{m'',k}|^4} \right)^{1/3}$$

and α_k^{Z3RO} is the per-user power normalization constant defined as

$$\alpha_k^{\text{Z3RO}} = \frac{\sqrt{M}}{\sqrt{\sum_{m'=M_s}^{M-1} |h_{m',k}|^2 + \gamma^2 \sum_{m=0}^{M_s-1} |h_{m,k}|^2}}.$$

The Z3RO precoder has a similar complexity as MRT with the main difference being the additional computation of γ_k.

2.2.3 PA Model

In this chapter, the Rapp PA model [41] is used. This model has a zero amplitude modulation to phase modulation (AM/PM) characteristic and an amplitude modulation to amplitude modulation (AM/AM) characteristic which is modeled by

$$y_m = \frac{x_m}{\left(1 + \left|\frac{x_m}{\sqrt{p_{\text{sat}}}}\right|^{2S}\right)^{\frac{1}{2S}}},$$

where the amplified signal at antenna m is denoted as y_m and S is the parameter which controls the smoothness of the transition from the linear region to the saturated region. $S = 2$ is chosen for the assessment performed in this study.

2.2.4 Channel Measurements

In [40], a measurement campaign was performed to capture real-world channel measurements. In this chapter, these open-source channel measurements[1] are used to evaluate the Z3RO precoder. The details of the measurements are summarized in Table 2.1. These channel measurements are available for 42 observer locations as illustrated in Fig. 2.2a. The set of locations is defined as $\mathcal{L}$. The measured complex channel gain from antenna m to an observer at location $l \in \mathcal{L}$ is denoted as $\tilde{h}_{m,l}$, while a complex channel gain from antenna m to user k is denoted as $h_{m,k}$. The observer location, i.e., unintended user location, is used to evaluate the distortion experienced at that specific position. This allows the study of the distortion level at both the user location and all unintended (observer) locations.

[1]The open-source measurements are available here: `https://dramco.be/massive-mimo/measurement-selector/#Sub-GHz`

Table 2.1: Overview of the open-source measurements.

Parameter	Symbol	Value
Carrier frequency	f_c	2.61 GHz
Number of base station antennas	M	32
Base station Array configuration		ULA
Type of BS antenna		Patch
Type of UE antenna		Dipole
BS height		7 m
UE height		1.5 m
Number of measurement positions		42

2.2.5 Received Signal

The received signal at observer location $l \in \mathcal{L}$ is given by

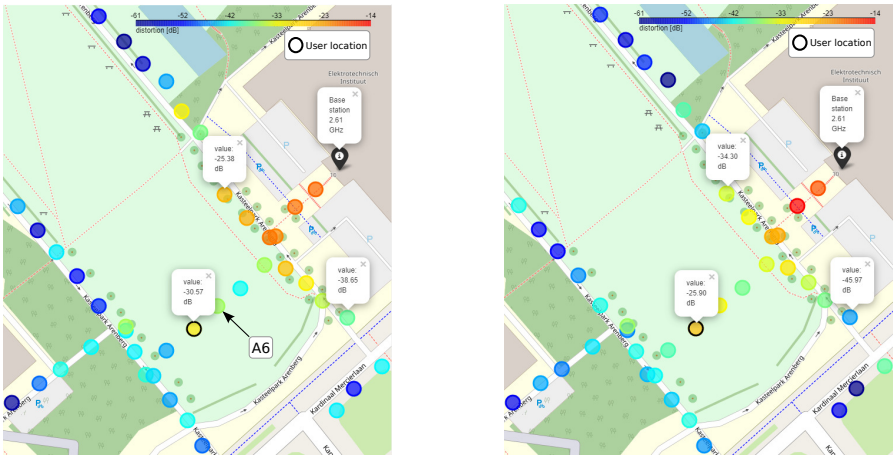
$$r_l = \sum_{m=0}^{M-1} \tilde{h}_{m,l} y_m + v_l,$$

where v_l is zero-mean complex Gaussian noise with variance σ_v^2 and y_m is the amplified transmit signal, precoded to target one or several users.

2.3 Performance Assessment

2.3.1 Performance Evaluation Metrics

In order to compare Z3RO precoding with MRT precoding, the absolute distortion level (expressed in dB), is evaluated. Given that the absolute distortion level is correlated with the OOB radiation, it is a measure of the potential OOB leakage, and thus the adjacent channel leakage ratio (ACLR). Consequently, studying the absolute distortion level allows us to get a better view on the OOB distortion level as well as the in-band (IB) distortion level (which impacts the SDR). The absolute distortion level at all observer locations can be evaluated using the following methodology. First, the received signal is decomposed in two parts: a desired part and an uncorrelated additive "noise" term following the Bussgang



(a) Z3RO (b) MRT

Figure 2.2: Heatmap of the distortion at all observer locations, for $p_{\text{in}}/p_{\text{sat}} = -3.1\,\text{dB}$. The distortion level at the user location is reduced for the Z3RO precoder, while it is slightly higher at unintended locations. The circles denote the measurement locations, and the black icon the base station. The position highlighted with an arrow is the position used in Fig. 2.4. (approximate area of $25\,000\,\text{m}^2$)

theorem [42, 43]. Thereafter, the signal-to-noise-and-distortion ratio (SNDR) and signal-to-noise-and-interference-and-distortion ratio (SNIDR) at the user location(s) are determined for the single and multi-user case.

2.3.2 Single-User Case

In this section, a scenario is considered where a single-user is served by the base station. Given that having more users spatially spreads out the nonlinear distortion, the single-user scenario represents the worst-case in terms of the coherent combining of the nonlinear distortion at the user location. We define the user location index as $\tilde{l} \in \mathcal{L}$.

Following the Bussgang decomposition [43], the received signal at location l for a single-user can be decomposed as

$$r_l = \underbrace{G_l s}_{\text{desired signal}} + \underbrace{d_l}_{\text{nonlinear distortion}} + \underbrace{v_l}_{\text{noise}}, \tag{2.1}$$

where $l \in \mathcal{L}$, d_l denotes the nonlinear distortion, that is uncorrelated with the transmit symbol s. G_l is the linear gain that can be calculated as $G_l = \mathbb{E}\left(r_l s^*\right)/p$ [43]. The received signal variance is denoted as $|G_l|^2 p$. The distortion variance is calculated using the knowledge that s, v and d_l are uncorrelated. This leads to the following expression,

$$\mathbb{E}\left(|d_l|^2\right) = \mathbb{E}\left(|r_l|^2\right) - |G_l|^2 p - \sigma_{v_l}^2. \tag{2.2}$$

Using this expression it is possible to calculate the SNDR at the user location for which index $l = \tilde{l}$,

$$\text{SNDR} = \frac{|G_{\tilde{l}}|^2 p}{\mathbb{E}\left(|d_{\tilde{l}}|^2\right) + \sigma_{v_{\tilde{l}}}^2}.$$

Next to the distortion level (2.2), the symbol rate at user k, expressed in bits/symbol is also considered. Assuming the worst-case of having the distortion Gaussian distributed, a lower bound on the achievable rate can be obtained as follows

$$R = \log_2(1 + \text{SNDR}). \tag{2.3}$$

Utilizing this methodology, the absolute distortion level, found in (2.2), is calculated at all observer locations. This is done at an input back-off of -3.1 dB, which is significantly deeper in saturation as compared to traditional systems. The number of saturated antennas for the Z3RO precoder is set to two ($M_s = 2$). As seen in Fig. 2.2, the distortion at the user location is reduced for the Z3RO precoder, which is to be expected given that the Z3RO precoder nulls the third-order distortion term at the user location. Additionally, the distortion is spread out spatially, which results in higher distortion values for the Z3RO precoder at the observer locations. This is consistent with the theoretical results obtained in [39], as can be seen in Fig. 2.3. The results illustrated in Fig. 2.2 solely depict one possible user location. However, given the 42 measurement locations, it is possible to place the user once at all of these 42 locations. For a thorough comparison of both precoders, the empirical cumulative distribution function (eCDF) when placing the user at every possible location is shown in Fig. 2.5a. This shows that the distortion level at the user location for the Z3RO precoder is significantly lower than for the MRT precoder. On average, the Z3RO precoder reduces the distortion at the user location by 6.03 dB. We can also observe that the overall distortion power for the Z3RO precoder is slightly higher compared to the MRT precoder at the unintended locations, i.e., the observer positions, confirming the theoretical behavior as derived in [39]. Furthermore, it also demonstrates that the worst-case scenario is avoided, i.e., for high distortion levels the Z3RO precoding

results in 5.76 dB less distortion than MRT. Besides the absolute distortion level, the symbol rate (R) is also considered. Fig. 2.4 shows the symbol rate when the noise variance (σ_v^2) is varied, given a constant $p_{\text{in}}/p_{\text{sat}} = -3.1\,\text{dB}$. As can be seen in Fig. 2.4, if the system is noise limited (i.e., $pM||h||^2/\sigma_v^2 < 20\,\text{dB}$), MRT shows a slightly larger symbol rate. However, when the system is distortion limited (i.e., $pM||h||^2/\sigma_v^2 > 20\,\text{dB}$), the Z3RO precoder has a higher symbol rate. This confirms and validates the intended behavior of the Z3RO precoder, as it improves performance when the system is distortion-limited.

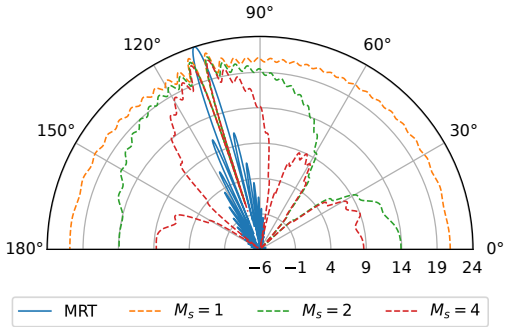


Figure 2.3: Third-order distortion pattern of MRT vs. Z3RO [dB], for $M = 32$ (adopted from [39]).

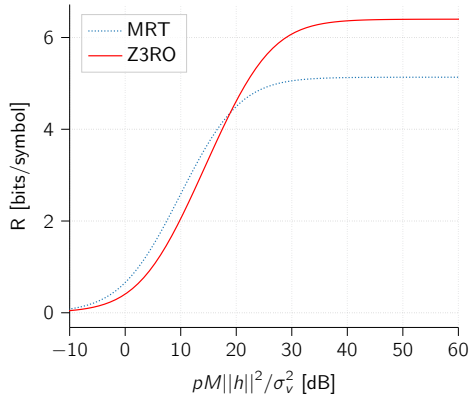
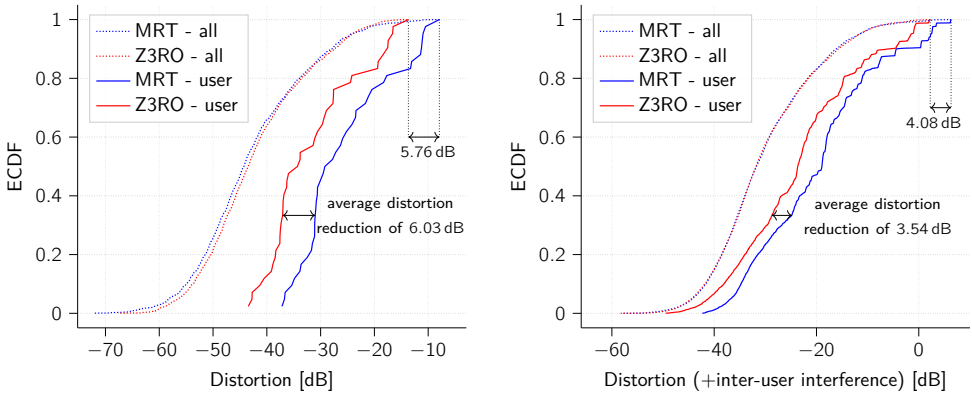


Figure 2.4: For a system that is noise limited, MRT has a higher symbol rate. For a distortion-limited system, Z3RO has a higher symbol rate. The considered user position is highlighted in Fig. 2.2. ($p_{\text{in}}/p_{\text{sat}} = -3.1\,\text{dB}$).



(a) Single-user case. ($M_s = 2$)

(b) Two-user case. ($M_s = 4$)

Figure 2.5: ECDF of the distortion level for the single and multi-user case. The distortion level at all possible user locations (solid) and at all user and observer locations (dashed) for MRT and Z3RO precoding. In both scenarios, the Z3RO precoder reduces both the distortion at the user location (solid) and the maximum experienced distortion at the observer locations (dashed). ($p_{\text{in}}/p_{\text{sat}} = -3.1\,\text{dB}$)

2.3.3 Two-User Case

For the two-user case the number of saturated antennas is set to four ($M_s = 4$). We define the user location index of user k as $\tilde{l}_k \in \mathcal{L}$. Both users use the same time and frequency resources. The same reasoning used in the single-user case can be followed for the multi-user case. The signal, intended for user k, received at observer l can be described as,

$$r_{l,k} = \underbrace{G_{l,k}s_k}_{\text{Desired signal}} + \underbrace{d_{l,k}}_{\substack{\text{Nonlinear distortion +} \\ \text{user interference}}} + \underbrace{v_l}_{\text{Noise}}, \tag{2.4}$$

where $l \in \mathcal{L}$, $k = 0, \ldots, K-1$. Note that now $d_{l,k}$ also includes inter-user interference. The Bussgang gain is computed as $G_{l,k} = \mathbb{E}\left(r_{l,k}s_k^*\right)/p_k$ [43]. The variance of the distortion and inter-user interference can be determined by,

$$\mathbb{E}\left(|d_{l,k}|^2\right) = \mathbb{E}\left(|r_{l,k}|^2\right) - |G_{l,k}|^2 p_k - \sigma_{v_l}^2. \tag{2.5}$$

The SNIDR at user k is then given by

$$\mathrm{SNIDR}_k = \frac{|G_{\tilde{l}_k,k}|^2 p_k}{\mathbb{E}\left(|d_{\tilde{l}_k,k}|^2\right) + \sigma^2_{v_{\tilde{l}_k}}}. \tag{2.6}$$

Equivalently to the single-user case, a lower bound on the achievable rate for the multi-user case can be obtained as follows,

$$R_k = \log_2(1 + \mathrm{SNIDR}_k). \tag{2.7}$$

The only difference in the multi-user case is that the achievable rate is also impacted by the inter-user interference.

Utilizing equation (2.5), Fig. 2.5b depicts the eCDF of the distortion level (now including the inter-user interference) when considering every possible two-user combination given the 42 measurement locations. This figure shows an average reduction of the distortion power at the user locations of 3.54 dB. The worst-case scenario, in terms of absolute distortion power, is also here reduced by 4.08 dB for Z3RO precoding. This demonstrates the ability of the Z3RO precoder to reduce the distortion level at the user location, even in the two-user case. We note that the distortion reduction is smaller than in the single-user case. However, this is to be expected given that the two-user case is less stringent as the distortion is more spatially spread out when more users are present.

2.4 Conclusion

In this chapter, the performance of the Z3RO precoder was assessed based on real-world channel measurements. The results demonstrate an average distortion reduction at the user location of 6.03 dB in the most stringent single-user case and 3.54 dB in the two-user case. Additionally, we showed that the lower-bound symbol rate for the Z3RO precoder is better as compared to MRT, when the system is distortion limited. Finally, we confirmed that the Z3RO precoder introduces slightly more distortion at the unintended locations. However, we see that the worst-case scenario is avoided, as for high distortion levels, the Z3RO precoded signals experience 5.76 dB less distortion than MRT, in the single-user case and 4.08 dB in the two-user case. This assessment validates the capability of the Z3RO precoder to operate PAs in massive MIMO systems closer to the saturation and hence increase the energy efficiency considerably. This study highlights the potential of distortion mitigation through distortion-aware precoding. However, currently, no inter-user interference cancellation is present in the Z3RO precoder, making it

challenging to deploy the precoder to scenarios with more users. Additionally, the Z3RO precoder is developed under the assumption of a third-order polynomial PA model, which is only valid as the PA enters saturation. When considering operating conditions where the PAs are deeply saturated, higher-order polynomial models need to be considered. Considering these extensions result in a complex non-convex optimization problem, which is non-trivial to solve with classical methods. The next chapter proposes a neural network-based precoder to perform distortion-aware precoding in the multi-user case under a higher-order polynomial PA model.

Chapter 3

Neural Network Approaches for Distortion-Aware Precoding

This chapter is based on the following papers:

> T. Feys, X. Mestre, and F. Rottenberg, "Self-Supervised Learning of Linear Precoders Under Non-Linear PA Distortion for Energy-Efficient Massive MIMO Systems," in *ICC 2023 - IEEE International Conference on Communications*, pp. 6367–6372, 2023 [27]
>
> T. Feys, L. Van der Perre, and F. Rottenberg, "Toward Energy-Efficient Massive MIMO: Graph Neural Network Precoding for Mitigating Non-Linear PA Distortion," *IEEE Transactions on Cognitive Communications and Networking*, vol. 11, no. 1, pp. 184–201, 2025 [28]

In the previous chapter, the problem of power amplifier (PA) nonlinearity in massive multiple input multiple output (MIMO) systems was introduced, and the performance of the closed-form Z3RO precoder was assessed. In this chapter, we propose the use of a graph neural network (GNN) to find a mapping from the channel matrix to the precoding matrix. This mapping is learned under the presence of non-linear PAs operating close to their saturation point, which increases their energy efficiency but, as a consequence, introduces non-linear distortion. The need for machine learning arises from the non-linear and non-convex nature of the optimization problem, which limits classical linear signal

processing solutions. Using a neural network (NN) allows to consider a more realistic scenario, i.e., serving multiple users and assuming a high-order polynomial PA model. Additionally, by learning the mapping from channel matrix to precoding matrix, a lot of the complexity is offloaded to the training step, which reduces the online computational complexity. This allows for a practical low-complexity solution even for the multi-user precoding problem under the presence of high-order non-linear PA distortion. First, it is studied which NN architectures are suitable for linear precoding, which is in line with research question Q2.1 (Which neural network architectures are suitable for linear precoding?). Next, it is shown that, in the distortion-limited regime, a GNN-based precoder achieves a higher sum rate as compared to zero forcing (ZF), ZF plus digital pre-distortion (DPD), and the distortion-aware beamforming (DAB) precoder from the state-of-the-art. The GNN achieves these gains by transmitting the non-linear distortion in non-user directions. Next to this, for a fixed sum rate, the achievable reduction in power consumption is computed. For this, both the power consumption reduction of the PAs and the power consumption increase of the digital signal processing related to the GNN are considered. This chapter addresses research questions Q2 'Can NN-based precoding be leveraged to allow PAs to operate close to saturation, while maintaining the data rate?' and Q4.1 'How much can the PA consumed power be reduced, and does this outweigh the additional inference power consumption of the required NN?'.

The remainder of this chapter is outlined as follows:

- Section 3.1 introduces the problem and gives an overview of the current state-of-the-art.
- Section 3.2 describes the system model along with the non-linear PA models which are used to train and evaluate the proposed GNN precoder. Next to this, the optimization problem considering high-order PAs is formulated. It is outlined how this formulation is used as a loss function to train a neural network precoder.
- We compare a number of neural network architectures for linear precoding in Section 3.3.1. This is done under the assumption of a simplified third-order PA model. A GNN is shown to outperform others such as circular convolutional neural networks (CCNNs) and multilayer perceptrons (MLPs). Based on this comparison, the GNN is selected for further training and evaluation on a high-order PA model.
- In Section 3.3.2, we characterize the GNN architecture and how it can be adopted for precoding. This architecture is further extended in Section 3.3.3

to include the signal-to-noise ratio (SNR) as an additional input, which allows the precoder to perform well over the full SNR range, rather than having to be retrained at each SNR point to reach the desired performance.

- The complexity of the proposed GNN precoder is analysed in Section 3.4. Additionally, the complexity of the benchmark DAB precoder from [44] has been computed and is used to compare with the GNN precoder.
- Section 3.5 provides extensive simulation results to validate the performance of the GNN-based precoder under the presence of high-order non-linear PAs. This includes the computation of the power consumption reduction, where the total consumed power (PA and processing) of the GNN precoder is quantified.
- Section 3.6 concludes the chapter.

3.1 Introduction

3.1.1 Problem Statement

We present an approach to improve the energy efficiency of PAs by operating them closer to saturation. To do so, we study how a neural network can learn a massive MIMO precoder that maximizes the sum rate, in the presence of non-linear distortion. This approach achieves a high capacity while using less back-off, thereby improving the energy efficiency. This, in turn, can reduce the operating expenses of wireless networks. Moreover, the cost of the PAs is an important component in the bill of materials (BOM) of wireless access equipment [7]. This cost is highly related to the maximum output power. Hence, by operating the PAs closer to saturation a lower maximum output power can be used to achieve the same average output power. The proposed method can be of particular interest for millimeter-wave systems which are typically highly non-linear [45, 46].

3.1.2 State-of-the-Art

The energy efficiency of the PA is typically limited by the need for a large back-off to remain within its linear regime. This is becoming less preferable given the growing emphasis on reducing energy consumption. Efforts to linearize the PA such as digital pre-distortion (DPD) are used in practical systems [47]. However, in massive MIMO systems, DPD has to be deployed at each antenna.

Recent studies show that when 5G systems use fully digital beamforming, a large number of antennas and a large bandwidth, these DPDs would introduce a power consumption that is even higher than the PA consumption itself [47]. Moreover, the performance of DPD is limited by clipping, i.e., the PA can only be linearized up to the saturation point, so that a relatively large back-off is still required, even for a perfect DPD. Peak-to-average power ratio (PAPR) reduction techniques have been studied, which can reduce the amount of back-off needed [48]. For instance, in [49] a constant-envelop precoder has been proposed. However, practical implementations encountered challenges due to the substantial complexity of digital signal processing.

Various other approaches to the distortion problem have been proposed. For instance, in [50] a neural network-based digital predistorter is proposed. This DPD is applied before the precoding stage which leads to a complexity that scales with the number of users, rather than the number of antennas. Similarly, in [51] an end-to-end system is proposed where neural-network-based precoding and decoding are jointly learned. In this scheme, the transmit symbol is essentially pre-distorted using a neural network, after which a zero-forcing precoder is applied, finally a neural network-based decoder is deployed at the receiver side. Alternatively, in [52] joint precoding and power allocation is proposed to solve the sum rate maximization problem in the presence of non-linear PAs. The authors solve the optimization problem using an iterative approach that alternates between optimizing the precoding and power allocation matrices. In a similar vein, in [53] a multiplier punitive method based power allocation algorithm is proposed which maximizes the spectral efficiency with awareness of the non-linear amplifiers. Additionally, in [54] the problem of non-linear PAs in distributed networks is considered. The authors propose a constrained unsupervised learning approach to solve the sum rate maximization problem in the presence of non-linear PAs with a particular focus on reducing the communication overhead between the distributed access points. While the above solutions offer certain benefits, they often lack in practical deployability due to their prohibitive complexity. For instance, in [51, 54, 50] a generic neural network architecture (MLP) is used, which results in high complexity. While [53, 52] rely on iterative procedures, which are challenging for real-life latencies.

Another compelling direction aims at incorporating knowledge of the distortion into the precoder design [39, 55, 26, 44]. This allows for the spatial suppression of the distortion in the user directions, producing considerable gains over classical precoders. Moreover, this offers an advantage over DPD based solutions as the distortion created by saturation-based clipping can be radiated in non-user directions, which improves the signal quality. Unfortunately, these solutions are still limited in their practical implementation, either because of their computational

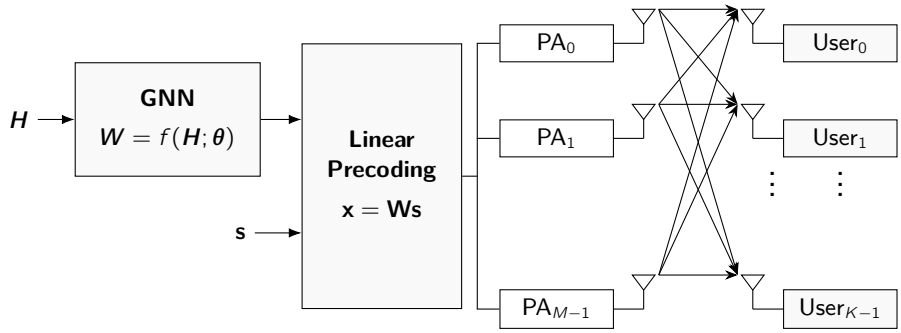


Figure 3.1: Massive MIMO downlink with neural network-based linear precoding and non-linear amplifiers at the transmitter.

complexity or in terms of simplifying assumptions that need to be addressed. More specifically, in [44], the solution to the precoding problem is obtained by solving a non-convex optimization problem with a projected gradient descent-based procedure. As the problem is non-convex, the procedure is executed multiple times to obtain a close-to-optimal solution, which requires a high computational complexity. Next to this, the zero third-order distortion (Z3RO) precoder as studied in Chapter 2 provides a globally optimal closed-form solution for the simplified single-user case and a third order PA model [39, 55]. There is thus a need for a solution that has low complexity and can address the challenging case of spatially multiplexing many users. Additionally, in order to work close to saturation, this solution should be designed for polynomial PA models that have a higher order than the third order model, as the third order model is only viable far away from saturation.

3.2 Problem Formulation

3.2.1 System Model

A downlink massive MIMO system with non-linear PAs and neural network-based linear precoding is considered as depicted in Figure 3.1. In this system, K single-antenna users are served by the base station (BS) using M transmit antennas. The complex symbol intended for user k is denoted as s_k, it is assumed to be zero mean circularly symmetric complex Gaussian with unit variance. This is a valid assumption for orthogonal frequency division multiplexing (OFDM) systems [56].

The symbols for different users are assumed to be uncorrelated. The linearly precoded symbol at antenna m is given by

$$x_m = \sum_{k=0}^{K-1} w_{m,k} s_k, \tag{3.1}$$

here $w_{m,k}$ is the precoding coefficient for user k at antenna m. In matrix form, the precoded symbol vector $\boldsymbol{x} \in \mathbb{C}^{M\times 1}$ is

$$\boldsymbol{x} = \boldsymbol{W}\boldsymbol{s} \tag{3.2}$$

where $\boldsymbol{W} \in \mathbb{C}^{M\times K}$ is the precoding matrix and $\boldsymbol{s} \in \mathbb{C}^{K\times 1}$ the symbol vector. The amplified transmit vector $\boldsymbol{y} \in \mathbb{C}^{M\times 1}$ is then given by

$$\boldsymbol{y} = \boldsymbol{\phi}(\boldsymbol{x}) \tag{3.3}$$

where $\boldsymbol{\phi}(\boldsymbol{x}) = [\phi(x_0), \cdots, \phi(x_{M-1})]^{\mathsf{T}}$ denotes the element-wise non-linear transformation caused by the PAs. Alternatively, in scalar notation the amplified signal at antenna m can be denoted as $y_m = \phi(x_m)$. The received signal vector $\boldsymbol{r} \in \mathbb{C}^{K\times 1}$ is

$$\boldsymbol{r} = \boldsymbol{H}^{\mathsf{T}}\boldsymbol{y} + \boldsymbol{v} = \boldsymbol{H}^{\mathsf{T}}\boldsymbol{\phi}(\boldsymbol{W}\boldsymbol{s}) + \boldsymbol{v}$$

with $\boldsymbol{H} \in \mathbb{C}^{M\times K}$ being the channel matrix. The vector $\boldsymbol{v} \in \mathbb{C}^{K\times 1}$ contains independently and identically distributed (i.i.d.) zero mean complex Gaussian noise samples with variance σ_v^2.

3.2.2 Non-Linear PA Models

The input signal of the PAs is considered to be a bandpass signal, which is described by the following complex baseband representation

$$x(t) = A(t)e^{j\varphi(t)}. \tag{3.4}$$

The output of a general non-linear PA with function $\phi(\cdot)$ is given by [57]

$$y(t) = \phi(x(t)) \tag{3.5}$$

$$= \phi_A\left(A(t)\right) e^{j\varphi(t)+\phi_\varphi(A(t))} \tag{3.6}$$

where $\phi_A(\cdot)$ and $\phi_\varphi(\cdot)$ represent the amplitude modulation to amplitude modulation (AM/AM) and amplitude modulation to phase modulation (AM/PM) transfer

function respectively. In this chapter, the non-linear PA is modeled as a complex-valued polynomial of order $2N+1$. For instance, an 11^{th} order PA gives $N=5$. Furthermore, only the odd order polynomial coefficients are considered given that they contribute to spectral components near the carrier frequency, while the even order coefficients cause spectral components at multiples of the carrier frequency [57]. Hence, the output of the PA at antenna m is given by [57, 58, 32]

$$\phi(x_m) = \sum_{n=0}^{N} \beta_{2n+1} x_m |x_m|^{2n} \tag{3.7}$$

$$= \beta_1 x_m + \beta_3 x_m |x_m|^2 + \cdots + \beta_{2N+1} x_m |x_m|^{2N} \tag{3.8}$$

where the polynomial coefficients β_{2n+1}, $0 \leq n \leq N$, are complex-valued coefficients that model both AM/AM and AM/PM distortion. Note that this formulation assumes that all M PAs have the same non-linear characteristic. Identical PA characteristics across the antenna array represent a worst-case in terms of coherent combining of the non-linear distortion [59]. Nevertheless, the extension to a system with different PA characteristics for each antenna can be made by replacing β_{2n+1} by $\beta_{2n+1}^{(m)}$, where the antenna index $(\cdot)^{(m)}$, is added to denote a unique complex coefficient per antenna m, as is done in [36].

In order to obtain proper polynomial coefficients at the desired input back-off (IBO), a least squares regression of the polynomial model to the modified Rapp model [60] is performed. The modified Rapp model has the same AM/AM distortion as the Rapp model used in Chapter 2. However, the modified Rapp model also includes an AM/PM distortion, both the AM/AM and AM/PM characteristics are described as

$$\phi_A(x_m) = \frac{|x_m|}{\left(1 + \left|\frac{x_m}{\sqrt{p_{\text{sat}}}}\right|^{2S}\right)^{\frac{1}{2S}}} \tag{3.9}$$

$$\phi_\varphi(x_m) = \frac{A|x_m|^\kappa}{1 + \left|\frac{x_m}{B}\right|^\kappa}. \tag{3.10}$$

The modified Rapp model coefficients are set as follows[1]: $S=2$, $\kappa=4$, $A=-0.315$, $B=1.137$ and the saturation power of the PA, p_{sat} is scaled in order to produce the desired IBO according to $\text{IBO} = p_{\text{in}}/p_{\text{sat}}$, with p_{in} being the average input power at each PA.

Additionally, as a benchmark a perfect DPD is considered. This is modeled as a soft limiter, i.e., a linear AM/AM characteristic up to a certain saturation

[1] Adapted from [58] to a PA with unit gain.

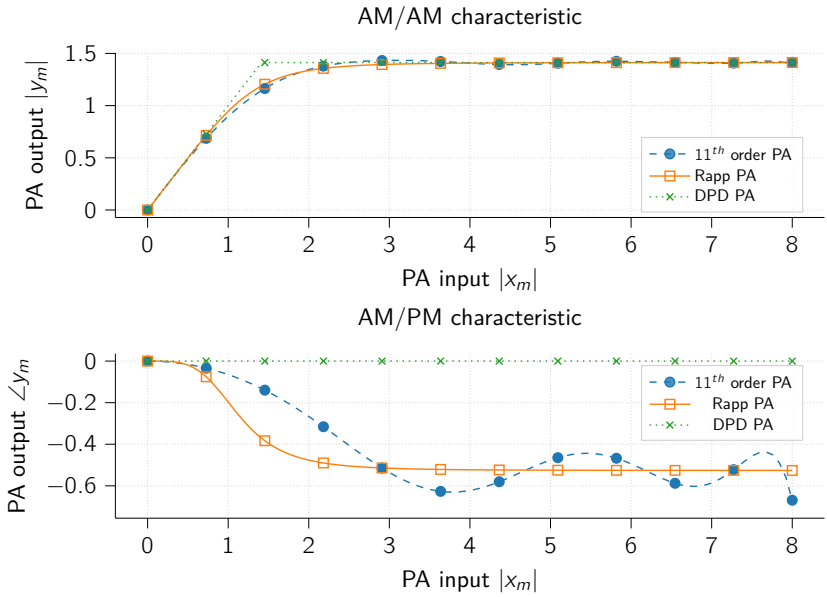


Figure 3.2: AM/AM and AM/PM characteristics of the considered power amplifier models: modified Rapp model, 11^{th} order polynomial model and a soft limiter/perfect DPD.

point where the output of the PA is clipped [57], which is equivalent to the AM/AM characteristic of the Rapp model for S going to infinity. The AM/AM characteristic is thus modeled as

$$\phi_A(x_m) = \begin{cases} |x_m| & \text{for} \quad |x_m| \leq \sqrt{p_{\text{sat}}} \\ \sqrt{p_{\text{sat}}} & \text{for} \quad |x_m| > \sqrt{p_{\text{sat}}} \end{cases}, \tag{3.11}$$

while the AM/PM conversion is zero. An overview of these PAs models is given in Figure 3.2.

3.2.3 Radiation Patterns

In Section 3.5, the radiation pattern of the proposed precoder is evaluated. For this, a pure line-of-sight (LoS) channel is considered. The channel from antenna m to user k can then be written as

$$h_{m,k} = e^{-jm\frac{2\pi}{\lambda_c} d\cos(\theta_k)}. \tag{3.12}$$

Here, $m\frac{2\pi}{\lambda_c}d\cos(\theta_k)$ represents the antenna-dependent phase shift when considering a uniform linear array (ULA) and a narrowband system. Additionally, λ_c denotes the carrier wavelength, θ_k the user angle and d the antenna spacing. The radiation pattern in an arbitrary direction $\tilde{\theta}$ can be obtained by

$$P(\tilde{\theta}) = \mathbb{E}\left(\left|\sum_{m=0}^{M-1} y_m e^{-jm\frac{2\pi}{\lambda_c}d\cos(\tilde{\theta})}\right|^2\right). \tag{3.13}$$

The radiation pattern of the linearly amplified signal is obtained by replacing y_m by x_m in (3.13) and is denoted as $P_{\text{lin}}(\tilde{\theta})$. The radiation pattern of the non-linear distortion is denoted as $P_{\text{dist}}(\tilde{\theta})$ and is obtained by replacing y_m by $\sum_{n=1}^{N}\beta_{2n+1}x_m|x_m|^{2n}$ in (3.13). Additionally, the signal-to-distortion ratio (SDR) radiation pattern is defined as $P_{\text{SDR}}(\tilde{\theta}) = \frac{P_{\text{lin}}(\tilde{\theta})}{P_{\text{dist}}(\tilde{\theta})}$.

3.2.4 Problem Formulation

The aim of this work is to find a precoding matrix that maximizes the sum rate, subject to a power constraint. An achievable sum rate R_{sum}, i.e., a lower bound on the capacity, is obtained by considering that the noise and distortion are jointly Gaussian distributed and independent from the data symbols, this can be seen as a worst case

$$R_{\text{sum}} = \sum_{k=0}^{K-1} \log_2(1 + \text{SNIDR}_k), \tag{3.14}$$

where SNIDR_k is the signal-to-noise-and-interference-and-distortion ratio (SNIDR) at user k. It can be computed based on the Bussgang decomposition [43], which implies that the received signal for user k can be written as $r_k = B_k s_k + d_k + v_k$. Here, d_k captures both the non-linear distortion and inter-user interference, which is uncorrelated to the transmit signal s_k and the noise v_k. The linear gain is given by $B_k = \mathbb{E}\left(r_k s_k^*\right)/p_k$, with $p_k = \mathbb{E}(s_k s_k^*)$ [43]. The received signal variance for user k is given by $|B_k|^2 p_k$. The distortion and inter-user interference can be computed as $\mathbb{E}\left(|d_k|^2\right) = \mathbb{E}\left(|r_k|^2\right) - |B_k|^2 p_k - \sigma_v^2$, given that d_k, s_k and v_k are uncorrelated. The SNIDR for user k is then given by

$$\text{SNIDR}_k = \frac{|B_k|^2 p_k}{\mathbb{E}\left(|d_k|^2\right) + \sigma_v^2}. \tag{3.15}$$

This general expression for the SNIDR can be evaluated numerically for a general PA model and is used for the evaluation in Section 3.5.

For training the NNs, the $(2N+1)^{\text{th}}$ order polynomial PA model in (3.7) is assumed, which leads to an analytical expression for the SNIDR [44, 45]. By applying Bussgang's theorem [43] to the amplification stage, we can write the amplified signal as

$$\boldsymbol{\phi}(\boldsymbol{x}) = \boldsymbol{G}\boldsymbol{x} + \boldsymbol{e}, \tag{3.16}$$

with $\boldsymbol{e} \in \mathbb{C}^{M\times 1}$ the non-linear distortion term and $\boldsymbol{G} \in \mathbb{C}^{M\times M}$ a diagonal matrix containing the Bussgang gains with the diagonal entries being $[\boldsymbol{G}]_{m,m} = \mathbb{E}[\phi(x_m)x_m^*]/\mathbb{E}[|x_m|^2]$.

When assuming the $(2N+1)^{\text{th}}$ order polynomial model and linear precoding ($\boldsymbol{x} = \boldsymbol{W}\boldsymbol{s}$), the gain matrix $\boldsymbol{G}$ can be written as a function of the precoding matrix $\boldsymbol{W}$ [45]

$$\boldsymbol{G}(\boldsymbol{W}) = \sum_{n=0}^{N} (n+1)!\beta_{2n+1}\boldsymbol{I}_M \text{diag}\left(\boldsymbol{C}_x\right)^n \tag{3.17}$$

The input covariance matrix is given by $\boldsymbol{C}_x = \mathbb{E}[\boldsymbol{x}\boldsymbol{x}^H] = \boldsymbol{W}\boldsymbol{W}^H$. From [44, 45], the covariance matrix of the non-linear distortion $\boldsymbol{e}$ can be derived as

$$\boldsymbol{C}_e(\boldsymbol{W}) = \sum_{n=1}^{N} \boldsymbol{L}_n \boldsymbol{C}_x \odot |\boldsymbol{C}_x|^{2n} \boldsymbol{L}_n^H, \tag{3.18}$$

with

$$\boldsymbol{L}_n = \frac{1}{\sqrt{n+1}} \sum_{l=n}^{N} \binom{l}{n} (l+1)!\beta_{2l+1}\boldsymbol{I}_M \text{diag}\left(\boldsymbol{C}_x\right)^{l-n}. \tag{3.19}$$

The received signal at user k can then be written as

$$r_k = \underbrace{\boldsymbol{h}_k^{\intercal}\boldsymbol{G}(\boldsymbol{W})\boldsymbol{w}_k s_k}_{\text{desired signal}} + \underbrace{\sum_{k'\neq k} \boldsymbol{h}_k^{\intercal}\boldsymbol{G}(\boldsymbol{W})\boldsymbol{w}_{k'} s_{k'}}_{\text{inter-user interference}} + \underbrace{\boldsymbol{h}_k^{\intercal}\boldsymbol{e}}_{\substack{\text{received} \\ \text{non-linear distortion}}} + \underbrace{v_k}_{\text{noise}}. \tag{3.20}$$

This leads to the following SNIDR expression for user k

$$\text{SNIDR}_k(\boldsymbol{W}) = \frac{|\boldsymbol{h}_k^{\intercal}\boldsymbol{G}(\boldsymbol{W})\boldsymbol{w}_k|^2}{\sum\limits_{k'\neq k} |\boldsymbol{h}_k^{\intercal}\boldsymbol{G}(\boldsymbol{W})\boldsymbol{w}_{k'}|^2 + \boldsymbol{h}_k^{\intercal}\boldsymbol{C}_e(\boldsymbol{W})\boldsymbol{h}_k^* + \sigma_v^2}. \tag{3.21}$$

Given this expression for the SNIDR, an achievable sum rate can be computed using (3.14). As such, the optimization problem we aim to solve can be formulated

as

$$\max_{\boldsymbol{W}} \quad R_{\text{sum}}(\boldsymbol{W}) \tag{3.22}$$

$$\text{s.t.} \quad \mathbb{E}\left(\sum_{m=0}^{M-1} |x_m|^2\right) = \text{Tr}\left(\boldsymbol{W}\boldsymbol{W}^H\right) \leq P_T,$$

where P_T is the total transmit power. The aim is thus to find a precoding matrix which maximizes the sum rate, subject to a power constraint[2], while the system is affected by PA nonlinearities.

3.2.5 Neural Network Training

The neural network-based precoder considered in this work takes as input the channel matrix $\boldsymbol{H}$ and outputs the linear precoding matrix $\hat{\boldsymbol{W}}$. This neural network can be represented as $\hat{\boldsymbol{W}} = f(\boldsymbol{H};\boldsymbol{\theta})$ i.e., a learned non-linear function mapping which is parameterized by $\boldsymbol{\theta}$. The channel matrix $\boldsymbol{H}$ is modeled as a random variable drawn from the considered channel distribution $p(\boldsymbol{H})$. Training is done in a self-supervised manner by maximizing the expected sum rate over the channel distribution, to obtain the neural network parameters

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \; \mathbb{E}_{\boldsymbol{H}\sim p(\boldsymbol{H})}[-R_{\text{sum}}(f(\boldsymbol{H};\boldsymbol{\theta}))]\,. \tag{3.23}$$

In practice, this expectation is intractable and is approximated by the empirical average over a training set $\mathcal{D} = \{\boldsymbol{H}^{(i)}\}_{i=0}^{N_{\text{train}}-1}$ of channel realizations sampled from $p(\boldsymbol{H})$, resulting in the empirical risk minimization problem

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \; -\frac{1}{N_{\text{train}}} \sum_{i=0}^{N_{\text{train}}-1} R_{\text{sum}}(f(\boldsymbol{H}^{(i)};\boldsymbol{\theta})). \tag{3.24}$$

Here the sum rate R_{sum} is given by (3.14), where the SNIDR is computed according to (3.21). Since the SNIDR in (3.21) directly depends on the output of the neural network (i.e., the precoding matrix $\hat{\boldsymbol{W}}$), the gradients of this loss function, with respect to the neural network parameters $\boldsymbol{\theta}$, can directly be computed using backpropagation. Given these gradients, the parameters of the NN are updated using the Adam optimizer [61], on mini-batches drawn from $\mathcal{D}$. Note that the average power constraint in (3.22) is implicitly taken into account in the loss

[2]For simplicity, the power constraint is taken before the PA, which neglects the non-linearly amplified power, which is small compared to the full transmit power. Due to the saturation characteristic of the PA, this can be seen as an upper bound on the actual transmit power.

function, as a power normalization layer at the output of the neural network enforces this constraint. The details of this normalization layer are given in Section 3.3.2.

3.2.6 Benchmark Algorithms

Throughout this chapter, the proposed solution is compared with a number of benchmark algorithms. First, maximum ratio transmission (MRT) and zero forcing (ZF) precoding are considered [19]

$$\boldsymbol{W}^{\text{MRT}} = \alpha \boldsymbol{H}^* \tag{3.25}$$

$$\boldsymbol{W}^{\text{ZF}} = \alpha \boldsymbol{H}^* (\boldsymbol{H}^{\intercal} \boldsymbol{H}^*)^{-1}, \tag{3.26}$$

where $\alpha = \sqrt{P_T / \text{Tr}(\boldsymbol{W}\boldsymbol{W}^H)}$ is a power normalization constant. The MRT and ZF precoding matrices are obtained under the assumption that the PA is linear. Hence, these two algorithms act as a benchmark in the linear regime where no distortion is present. Additionally, they represent a worst-case scenario, highlighting how the system's performance deteriorates when non-linear distortion is present and no mitigation techniques are adopted.

Second, the Z3RO precoder as studied in the previous chapter is considered [39]. The closed-form expression for the precoding coefficient at antenna m is given by

$$w_m^{\text{Z3RO},M_s} = \alpha h_m^* \begin{cases} -\gamma & \text{if } m = 0, \dots, M_s - 1 \\ 1 & \text{otherwise} \end{cases}, \tag{3.27}$$

here $\gamma = \left(\sum_{m'=M_s}^{M-1} |h_{m'}|^4 / \sum_{m''=0}^{M_s-1} |h_{m''}|^4 \right)^{1/3}$ is the gain of the saturated antennas and M_s is the number of saturated antennas, which is set to $M_s = 1$ in all simulations, as this gives the globally optimal solution [55]. In [55], this precoder was shown to be optimal in the single-user case and when a third-order PA model is assumed. Hence, in this chapter, it is used to evaluate how close the NN precoder is to the optimal solution, under these assumptions.

For the more general multi-user case and when high-order PAs are assumed, the problem becomes non-convex, consequently, no optimal solution is known. However, in [44] the non-convex optimization problem is approximately solved using an iterative projected gradient ascent method, which is denoted as the DAB precoder. For the multi-user and high-order PA case, this algorithm is used as a benchmark. The precoding matrix obtained by this algorithm is denoted as $\boldsymbol{W}^{\text{DAB}}$

and can be obtained by performing the following algorithm for I iterations

$$\boldsymbol{W}^{(i+1)} = \left[\boldsymbol{W}^{(i)} + \mu^i \nabla_{\boldsymbol{W}} R_{\text{sum}}\left(\boldsymbol{W}^{(i)}\right)\right]^{+}_{\mathbb{E}\|\boldsymbol{\phi}(\boldsymbol{W}\boldsymbol{s})\|^2 = P_T} \tag{3.28}$$

Here μ^i is the step size for the gradient update at iteration i and $[\cdot]^{+}_{\mathbb{E}\|\boldsymbol{\phi}(\boldsymbol{W}\boldsymbol{s})\|^2=P_T}$ denotes the projection of the obtained solution, such that the power constraint is met[3].

3.3 Neural Networks for Precoding

3.3.1 Suitable Neural Network Architectures

Neural networks, with 1 hidden layer and sufficient neurons, can approximate any continuous non-linear function with arbitrary accuracy [62]. Subsequently, they can be used to learn a mapping from channel matrix $\boldsymbol{H}$ to precoding matrix $\hat{\boldsymbol{W}}$. Unfortunately, the universal approximation theorem is not constructive, i.e., it does not provide a way to select the NN architecture and number of neurons that can achieve this arbitrary accuracy. A MLP is a general fully connected feedforward neural network, which is a universal function approximator [62, 63]. Given the universal function approximation theorem, the hypothesis space, that is the set of all possible functions the MLP can represent, is theoretically infinite when the network has sufficient neurons. As such, in theory, MLPs could be used to learn the mapping between channel and precoding matrix. In practice, however, there is a trade-off to be made. A large hypothesis space has the merit of being able to learn many complex functions. However, a large hypothesis space typically leads to a very large number of trainable parameters. Consequently, these MLPs are very hard to train, require a large amount of training data, and have a high computational complexity at inference time.

Hence, it is advantageous to reduce the size of this hypothesis space in an informed manner. This can be done in two ways. (1) By decreasing the size of the network (e.g., by using less neurons). (2) By imposing a certain structure (i.e., an inductive bias) on the NN architecture, which limits the types of functions that can be learned. A reduced hypothesis space typically corresponds to fewer trainable parameters, which means that less training data is needed and training becomes

[3]Note that in [44] the power constraint is considered after the PA in contrast to our approach where the power constraint is enforced before the PA. This implies that our proposed precoder abides by both power constraints, i.e., before and after the PA, while the DAB precoder only abides by the constraint after the PA.

easier. This, however, comes at the cost of a less expressive model, potentially leading to the fact that the desired function to be learned is no longer covered by this reduced hypothesis space [64]. However, if we can adequately reduce the size of the hypothesis space such that the desired function to be learned is still covered by the (reduced) hypothesis space of the NN, the reduction in expressivity of the model does not affect the learning performance. Hence, in this scenario, only the merits of reducing the hypothesis space remain, namely, more scalable NNs, with fewer learnable parameters, that are easier to train and have a lower computational complexity [65].

More formally, this can be understood through statistical learning theory [66]. We follow the reasoning from [64] to formally explain how the reduction of the hypothesis space can lead to improved learning performance. First, we define the risk of a model f as the expectation of the loss function, which in our case is the negative sum rate

$$\mathcal{R}(f) = \mathbb{E}\left(-R_{\text{sum}}\left(f(\mathbf{H};\boldsymbol{\theta})\right)\right). \tag{3.29}$$

The goal when training a neural network is to minimize the risk. Next, we define the hypothesis space $\mathcal{F}$ of a neural network as the space of all functions that can be learned by the NN. Statistical learning theory defines the generalization error of model $\hat{f}$ (i.e., the learning performance) as the sum of the approximation and estimation errors.

$$\underbrace{\mathcal{R}(\hat{f}) - \mathcal{R}^{\star}}_{\text{generalization error}} = \underbrace{[\mathcal{R}(\hat{f}) - \mathcal{R}(f^{*})]}_{\text{estimation error}} + \underbrace{[\mathcal{R}(f^{*}) - \mathcal{R}^{\star}]}_{\text{approximation error}}. \tag{3.30}$$

Here $\mathcal{R}^{\star}$ is the minimal achievable risk associated with the optimal function $f^{\star}$. $\hat{f}$ is the learned model and $f^{*} = \text{argmin}_{f\in\mathcal{F}}\mathcal{R}(f)$ is the best model within the hypothesis space $\mathcal{F}$. The estimation error represents the difference between the risk of the learned function and the risk of the best possible function in the hypothesis space $\mathcal{F}$, hence the estimation error is caused by the limited size of the training set and the limitations of the optimization of the learned weights. On the other hand, the approximation error represents the difference between the risk of the best possible function in $\mathcal{F}$ and the absolute minimal achievable risk considering all possible functions, i.e., not limited by $\mathcal{F}$. Hence the approximation error is determined by the selection of the hypothesis space $\mathcal{F}$.

As discussed above, the size of the hypothesis space can be adjusted in two ways. When the size of $\mathcal{F}$ is increased, for a fixed amount of training data, the NN becomes more expressive, and hence the approximation error will decrease, on the other hand, the estimation error will increase as more parameters need to be learned with the same amount of training data. This trade-off justifies why

fewer trainable parameters can lead to lower or higher performance. Generally, if the hypothesis space decreases the approximation error increases. However, if the hypothesis space is carefully reduced from $\mathcal{F}$ to $\mathcal{F}'$, such that the best model within $\mathcal{F}$, f^* is still covered by $\mathcal{F}'$, the approximation error will remain the same while the estimation error is reduced.

The remaining question is how to adequately reduce the hypothesis space for the precoding problem without decreasing the learning performance. If we follow method (1) and select a general MLP and reduce its number of neurons, the estimation error will decrease, however, the approximation error increases, which is undesirable. Note that despite this, the size of the NN does need to be selected, this was done empirically by training NNs of different sizes and evaluating their performance on a validation set. Alternatively, we can follow method (2) to impose a certain structure on the NN. One can note that the ordering of the antennas and users in the channel and precoding matrices ($\boldsymbol{H}$, $\boldsymbol{W}$), is arbitrary. In other words, if the order of the users or antennas in the channel matrix $\boldsymbol{H}$ is permuted, this does not affect the physical underlying system, only its mathematical representation. Consequently, the order of the precoding vectors in $\boldsymbol{W}$ should be permuted accordingly, but the sum rate remains unchanged. This property is known as permutation equivariance [67]. More formally, a precoding function $\boldsymbol{W} = f(\boldsymbol{H})$ is permutation equivariant if the following holds $f(\boldsymbol{\Pi}_1 \boldsymbol{H} \boldsymbol{\Pi}_2^\intercal) = \boldsymbol{\Pi}_1 f(\boldsymbol{H}) \boldsymbol{\Pi}_2^\intercal$, $\forall\ \boldsymbol{\Pi}_1, \boldsymbol{\Pi}_2 \in \mathbb{P}$, where $\boldsymbol{\Pi}_1, \boldsymbol{\Pi}_2$ are permutation matrices and $\mathbb{P}$ is the set of all possible permutation matrices [67]. Given that all precoding functions are permutation equivariant, this can be leveraged to reduce the hypothesis space of the desired neural network, as it should only be able to learn permutation equivariant functions. Hence, when a NN architecture naturally exhibits this permutation equivariance property, it reduces the size of the hypothesis space, which reduces the estimation error, but the approximation error is unchanged as the optimal function is still within this reduced hypothesis space [64].

CCNNs are shift equivariant [67], which is a more general property then permutation equivariance. Hence their hypothesis space covers all permutation equivariant functions but is larger than the space of solely permutation equivariant functions. More formally this implies that $f_{\text{CCNN}}(\boldsymbol{S}_1 \boldsymbol{H} \boldsymbol{S}_2^\intercal) = \boldsymbol{S}_1 f_{\text{CCNN}}(\boldsymbol{H}) \boldsymbol{S}_2^\intercal$, $\forall\ \boldsymbol{S}_1, \boldsymbol{S}_2 \in \mathbb{S}$, where $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are circulant permutation matrices that represent the shift operation and $\mathbb{S}$ is the set of all circulant permutation matrices [67]. Given the additional circulant nature of these permutation matrices $\mathbb{S} \subseteq \mathbb{P}$. Permutation equivariant functions are more constrained in comparison to shift equivariant functions. Given the fact that every permutation equivariant function is also shift equivariant, it is clear that the hypothesis space of shift equivariant functions is larger than the hypothesis space of permutation equivariant functions. In [27], we showed that these CCNNs can be used to learn precoding functions that can

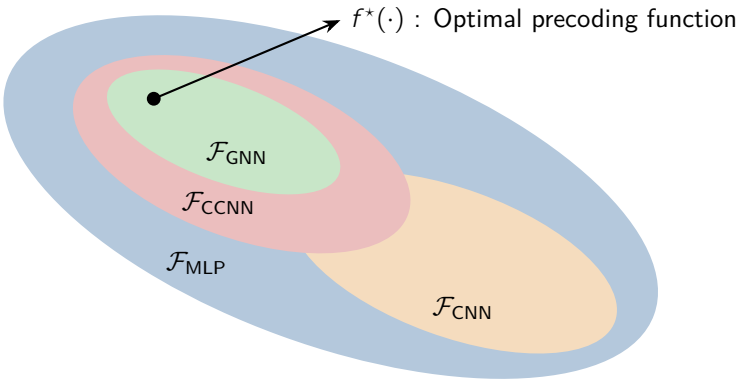
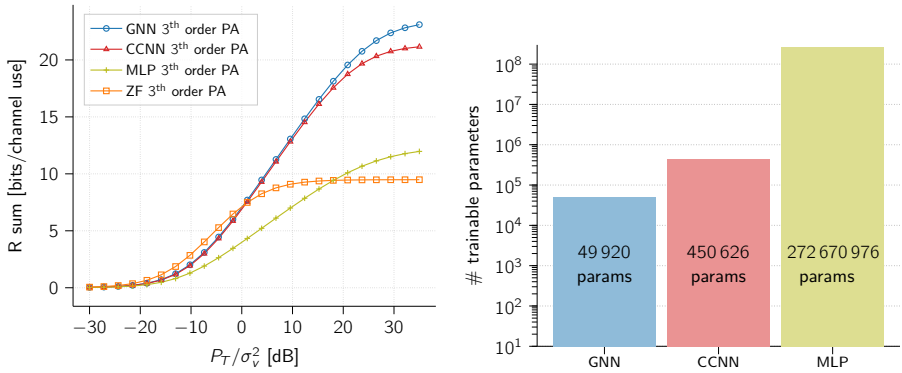


Figure 3.3: Hypothesis spaces of different neural network architectures, assuming infinite-capacity models.

cancel third-order non-linear distortion. In the current study, we show that GNNs can produce even better results, in the presence of both third and higher order non-linear distortion. GNNs are naturally permutation equivariant [68], which gives them the appropriate inductive bias for precoding.

The hypothesis spaces of the different NN architectures are illustrated in Figure 3.3. Note that for clarity Figure 3.3 assumes that the approximation error for MLPs, CCNNs and GNNs is zero, as the optimal function $f^\star$ lies within the hypothesis space of the MLP, CCNN and GNN ($\mathcal{F}_{\text{MLP}}$, $\mathcal{F}_{\text{CCNN}}$ and $\mathcal{F}_{\text{GNN}}$). This is only the case in the asymptotic sense, given the universal approximation theorems for MLPs and GNNs, the error tolerance on the function approximation approaches zero $\epsilon \to 0$ as the number of hidden units goes to infinity $N_{\text{hidden}} \to \infty$ [62, 63]. However, for the purpose of this conceptual comparison, we assume infinite-capacity models and focus on the architectural inductive bias. Consequently, we consider this asymptotic regime in which all model classes have sufficient capacity to approximate the optimal precoding function $f^\star$. Since $f^\star$ is permutation equivariant, MLPs, CCNNs, and GNNs are universal approximators for this function class, and the approximation error is assumed to be zero. In conclusion, all three architectures are expressive enough to represent $f^\star$, but some do so within a smaller, more structured hypothesis space, while traditional convolutional neural networks (CNNs) have a clear mismatch in terms of their hypothesis space, leading to a significant approximation error.

To illustrate this experimentally, Figure 3.4a denotes the performance of the different NN architectures for a scenario where $M = 64$, $K = 2$ and third-order



(a) (b)

Figure 3.4: (a) Comparison of an MLP, CCNN from [27] and GNN for $M = 64$, $K = 2$. Trained and evaluated on a third-order non-linear PA at -3 dB IBO, with $\beta_3 = -0.07781605 - 0.0401193j$. (200K training samples, 10K test samples) (b) Comparison of the number of trainable parameters of these three architectures.

non-linear amplifiers are considered. The dataset used in this experiment is partitioned into a training set of 200,000 channel realizations, a validation set of 2,000, and a test set of 10,000. of All 3 architectures are matched in terms of their representation capacity i.e., each of them has 6 layers and 64 features[4] per layer. From this image it is clear that the GNN achieves the highest sum rate as compared to the CCNN and MLP. The CCNN achieves a slightly lower sum rate, because of the shift equivariance it captures, which is still a relatively good inductive bias for precoding. The MLP has a very general structure and can barely produce a sum rate that is higher than that of the ZF precoder. Moreover, in Figure 3.4b the number of learnable parameters of each NN architecture is illustrated. This again highlights the proficiency of the GNN as it only has 50 000 learnable parameters as compared to the 450 000 and 270 000 000 of the CCNN and MLP.

3.3.2 Graph Neural Network Precoding

GNNs generally come in two forms, namely GNNs that learn a hidden representation of the nodes in a graph and GNNs that learn a hidden representation of the edges in a graph. For the precoding problem, we focus on GNNs that learn representations

[4] 64 features indicates that the real and imaginary part of each precoding coefficient is represented by a feature vector of dimension 64, resulting in a total of $64MK2$ features.

of the edges of the graph. In general, the goal is thus to start from a number of input edge features and learn edge embeddings. Such a NN consists of a number of layers or message-passing iterations, where each layer/iteration produces a new representation/embedding of the edges. An embedding/representation of edge (a, b) in layer l of a GNN is expressed as $\boldsymbol{z}_{(a,b)}^{(l)}$. A typical GNN layer consists of two steps, namely, a message-passing/aggregation step and an update step. Hence, a single layer for edge (a, b) can be expressed as

$$\boldsymbol{z}_{(a,b)}^{(l+1)} = \text{UPDATE}^{(l)}\left(\boldsymbol{z}_{(a,b)}^{(l)}, \boldsymbol{m}_{\mathcal{N}(a)}^{(l)}, \boldsymbol{m}_{\mathcal{N}(b)}^{(l)}\right) \tag{3.31}$$

Here $\boldsymbol{m}_{\mathcal{N}(a)}, \boldsymbol{m}_{\mathcal{N}(b)}$ are 'messages' that are aggregated from the neighboring edges of the two nodes connected to edge (a, b), these messages are defined as

$$\boldsymbol{m}_{\mathcal{N}(a)}^{(l)} = \text{AGGREGATE}^{(l)}\left(\{\boldsymbol{z}_{(a,x)}^{(l)}, \forall x \in \mathcal{N}(a)\}\right) \tag{3.32}$$

$$\boldsymbol{m}_{\mathcal{N}(b)}^{(l)} = \text{AGGREGATE}^{(l)}\left(\{\boldsymbol{z}_{(b,x)}^{(l)}, \forall x \in \mathcal{N}(b)\}\right). \tag{3.33}$$

Here 'UPDATE' and 'AGGREGATE' are differentiable functions. For a comprehensive overview of commonly used update and aggregate functions we refer to [68].

More specifically, the precoding problem can be represented as learning a mapping from a $M \times K$ MIMO channel matrix to a $M \times K$ precoding matrix over a graph, a visualization of this is given in Figure 3.5. The graph can be represented as follows: each antenna m is denoted by a node, and each user k is denoted by a node. The edges between the antennas and users represent the wireless channel and have as input features the channel coefficients. This graph is defined as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \mathcal{V}_M \cup \mathcal{V}_K$ is the set of nodes, with $\mathcal{V}_M$ the antenna nodes and $\mathcal{V}_K$ the user nodes, and $\mathcal{E}$ the set of edges, which contains all edges between the antenna and user nodes i.e., $(m, k) \in \mathcal{E}\ \forall m \in \mathcal{V}_M, k \in \mathcal{V}_K$. This graph is undirected, meaning that $(m, k) \in \mathcal{E} \leftrightarrow (k, m) \in \mathcal{E}$, as is the physical channel between each user and BS antenna. For this graph, a single GNN layer that updates the representation of edge (m, k) can be defined as

$$\boldsymbol{z}_{(m,k)}^{(l+1)} = \text{UPDATE}_{\text{edge}}\left(\boldsymbol{z}_{(m,k)}^{(l)}, \boldsymbol{m}_{\mathcal{N}(m)}^{(l)}, \boldsymbol{m}_{\mathcal{N}(k)}^{(l)}\right) \tag{3.34}$$

$$= \sigma\left(\boldsymbol{W}_{\text{edge}}^{(l)}\boldsymbol{z}_{(m,k)}^{(l)} + \boldsymbol{W}_{m}^{(l)}\boldsymbol{m}_{\mathcal{N}(m)}^{(l)} + \boldsymbol{W}_{k}^{(l)}\boldsymbol{m}_{\mathcal{N}(k)}^{(l)}\right), \tag{3.35}$$

where $\boldsymbol{z}_{(m,k)}^{(l+1)}$ denotes the hidden representation of the edge (m, k) at layer/iteration $l + 1$ of the GNN, $\sigma(\cdot)$ is a non-linear activation function and the matrices $\boldsymbol{W}_{\text{edge}}^{(l)}, \boldsymbol{W}_{m}^{(l)}, \boldsymbol{W}_{k}^{(l)} \in \mathbb{R}^{d_{l+1} \times d_l}$ are learnt using a form of stochastic gradient descent.

Note that the dimensions of these matrices, d_{l+1} and d_l, determine the number of features in layer $l+1$ and l respectively. The aggregation operations are illustrated in Figure 3.5a and 3.5b and are defined as follows

$$\boldsymbol{m}^{(l)}_{\mathcal{N}(m)} = \frac{1}{|\mathcal{N}(m)|} \sum_{k' \in \mathcal{N}(m)} \boldsymbol{z}^{(l)}_{(m,k')} \tag{3.36}$$

$$\boldsymbol{m}^{(l)}_{\mathcal{N}(k)} = \frac{1}{|\mathcal{N}(k)|} \sum_{m' \in \mathcal{N}(k)} \boldsymbol{z}^{(l)}_{(m',k)}. \tag{3.37}$$

The input and output for edge (m, k) of the network are defined in the following way

$$\mathbf{z}^{(0)}_{(m,k)} = \left[\Re\{[\mathbf{H}]_{m,k}\}, \Im\{[\mathbf{H}]_{m,k}\}\right]^{\mathsf{T}} \tag{3.38}$$

$$\mathbf{z}^{(L-1)}_{(m,k)} = \left[\Re\{[\hat{\mathbf{W}}]_{m,k}\}, \Im\{[\hat{\mathbf{W}}]_{m,k}\}\right]^{\mathsf{T}}. \tag{3.39}$$

In order to respect the power constraint, a power normalization layer is added, which consists of a scalar normalization given by

$$\hat{\boldsymbol{W}}^{norm} = \alpha \hat{\boldsymbol{W}}, \tag{3.40}$$

with $\alpha = \sqrt{P_T / \mathrm{Tr}(\hat{\boldsymbol{W}}\hat{\boldsymbol{W}}^H)}$.

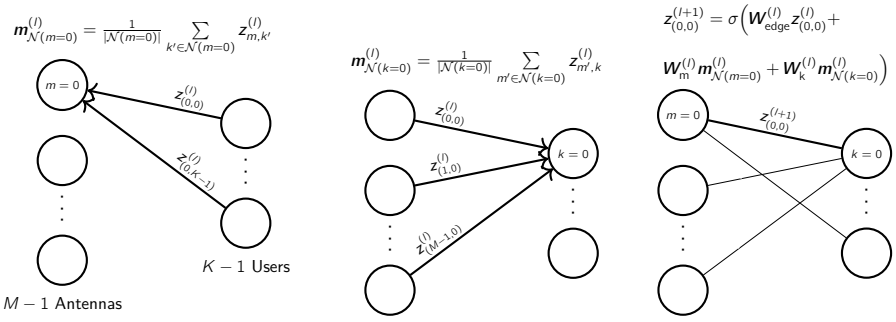
3.3.3 Leveraging SNR Regime Information

For the neural network architecture as described in Section 3.3.2, the network is typically trained at a fixed SNR ($= P_T/\sigma_\nu^2$) point. This results in the fact that the performance of the neural network is best close to the SNR value it is trained at. In this section, it is outlined how the SNR can be included as an input edge feature. This allows the neural network to be trained at a number of SNR points, depending on the input the neural network can differentiate in which SNR regime it is operating and deliver the best performance. To do so the input to the GNN is defined in the following way

$$\mathbf{z}^{(0)}_{(m,k)} = \left[\Re\{[\mathbf{H}]_{m,k}\},\ \Im\{[\mathbf{H}]_{m,k}\},\ \mathrm{SNR}_{\mathrm{norm}}\right]^{\mathsf{T}}. \tag{3.41}$$

To ensure stability during training we include the normalized SNR as input rather than the absolute SNR. The normalized SNR is defined as follows

$$\mathrm{SNR}_{\mathrm{norm}} = \frac{1}{\mathrm{SNR}_{\max}} \frac{P_t}{\sigma_\nu^2}. \tag{3.42}$$



(a) Message passing from users to antennas in layer l for edge $(0, 0)$.

(b) Message passing from antennas to user in layer l for edge $(0, 0)$.

(c) Update the edge feature using learned weight matrices for edge $(0, 0)$.

Figure 3.5: GNN visualized for a single edge $(0, 0)$ in the graph. In (a) and (b) message passing is performed. Note that in the first layer $l = 0$ the channel coefficients act as input edge features i.e., $\mathbf{z}_{(m,k)}^{(0)} = \left[\Re\{h_{m,k}\}, \Im\{h_{m,k}\}\right]^{\mathsf{T}}$. In (c) the messages and edge features are used to update the edge feature which acts as the output of this layer. This is repeated for each layer to generate the outputs of the network. The outputs of the network are the edge features in the final layer which represent the precoding coefficients $\mathbf{z}_{(m,k)}^{(L-1)} = \left[\Re\{w_{m,k}\}, \Im\{w_{m,k}\}\right]^{\mathsf{T}}$. (Note that the power normalization layer is omitted for simplicity.)

Training is done for an SNR ($= P_T/\sigma_\nu^2$) range of -30 to 30 dB, with a stepsize of 5 dB, hence $\text{SNR}_{\text{max}} = 30$ dB. During training, the SNR for each training example is randomly selected from this range. During testing, the sum rate is evaluated for an SNR range of -30 to 30 dB, however this time with a spacing of 2.5 dB. This ensures that SNR points for which the model is not trained are also tested.

3.4 Complexity and Real-Time Analysis

In this section, the complexity of the proposed solution is quantified and compared against the benchmark precoders. Next to this, it is outlined which operations of the GNN can be performed in parallel in order to produce faster execution times for real-time implementations. In order to quantify the computational complexity of the proposed solution, the number of real floating-point additions and multiplications are computed. Note that addition and multiplication carry a different computational cost, with multiplication being the most costly one.

However, this highly depends on the hardware platform [69]. Hence, both the number of real floating-point additions and multiplications are computed. Additionally, the number of floating point operations (FLOPs) is computed. A FLOP is defined as either a real floating point multiplication or addition. When considering complex multiplications or additions a conversion to real multiplications and additions is made. A complex multiplication requires 4 real multiplications and 2 real additions, while a complex addition requires 2 real additions. Note that the number of computations needed to perform the power normalization of the precoders is neglected, as this is equal across the precoders[5]. An overview of the complexity analysis can be found in Table 3.1.

In order to perform one forward pass of the GNN (3.35) has to be computed L times, namely the number of layers in the GNN, for each edge $(m, k) \in \mathcal{E}$ i.e., MK times. Note that our GNN architecture has $L = 8$ layers, meaning it consists out of an input layer, 6 hidden layers and 1 output layer. The number of in- and output features is fixed to $d_0 = d_{L-1} = 2$ and the number of features in each hidden layer is constant at $d_l = 128\ \forall l \in \{2, \cdots, L-2\}$. The GNN first performs message passing as illustrated in Figure 3.5a and 3.5b i.e., computing $\boldsymbol{m}^{(l)}_{\mathcal{N}(m)}$ and $\boldsymbol{m}^{(l)}_{\mathcal{N}(k)}$, which can be done by computing (3.36) and (3.37) this has the following complexity:

- Input layer: $2(M-1) + 2(K-1)$ real floating point additions
- Output and Hidden layers: $(L-1)(d_l(M-1) + d_l(K-1))$ real floating point additions

Subsequently, the update step is performed as illustrated in Figure 3.5c. This involves a number of matrix multiplications with the learned weight matrices, followed by the addition of these products, this has the following complexity:

- Input and output layer: $12d_l$ real multiplications and $11d_l - 2$ real additions
- Hidden layers: $(L-2)3d_l^2$ real multiplications and $(L-2)(3d_l^2 - d_l)$ real additions

Note that the complexity of the non-linear activation function is neglected as it is often approximated based on lookup tables, which requires few computations [70].

[5]Note that the DAB precoder employs a more complex power normalization which is performed after the amplification stage. However, a simpler power normalization before the PA could be used for this precoder, which would render the complexity of the normalization across all precoders equal. As such we decided not to include this term into the complexity analysis.

In total the GNN needs the following number of real floating point multiplications

$$\underbrace{MK}_{\text{edges}}\Bigg(\underbrace{\overbrace{6d_l}^{\text{(a)}}}_{\text{input layer}}+\underbrace{(L-2)\overbrace{3d_l^2}^{\text{(a)}}}_{\text{hidden layers}}+\underbrace{\overbrace{6d_l}^{\text{(a)}}}_{\text{ouput layer}}\Bigg) \tag{3.43}$$

and the following number of real floating point additions

$$\begin{aligned}&\underbrace{MK}_{\text{edges}}\Bigg(\underbrace{\overbrace{5d_l}^{\text{(a)}}+\overbrace{2(M-1+K-1)}^{\text{(b)}}}_{\text{input layer}}+\\&\underbrace{(L-2)\Big(\overbrace{3d_l^2-d_l}^{\text{(a)}}+\overbrace{d_l(M-1+K-1)}^{\text{(b)}}\Big)}_{\text{hidden layers}}+\underbrace{\overbrace{6d_l-2)}^{\text{(a)}}}_{\text{output layer}}\Bigg)\end{aligned} \tag{3.44}$$

where the term (a) denotes the computations needed to perform the matrix multiplication with the learned weight matrices followed by the addition of the resulting products and (b) denotes the computations needed for the message passing. This scales as $\mathcal{O}\left(MKd_l^2L\right)$ floating point multiplications and $\mathcal{O}\left(MKd_l^2L+M^2Kd_lL+MK^2d_lL\right)$ floating point additions. In terms of FLOPs this corresponds to $\mathcal{O}\left(MKd_l^2L+M^2Kd_lL+MK^2d_lL\right)$. Ultimately, it is important to highlight that the utilization of NNs offers a distinct advantage. Namely, a substantial portion of the computational tasks can be effectively parallelized, thereby resulting in a faster execution time. In our specific scenario, the execution of (3.35) can be parallelized across all edges, yielding a reduced number of FLOPs that need to be executed serially, namely $\mathcal{O}\left(d_l^2L+Md_lL+Kd_lL\right)$. This can be leveraged to achieve real-time operation.

3.4.1 Complexity of Zero Forcing Precoder

In order to compute the ZF precoder as given in (3.26), first a matrix multiplication of a $K\times M$ with an $M\times K$ matrix is computed, which requires MK^2 complex multiplications and $K^2(M-1)$ complex additions, this corresponds to $2MK^2-K^2$ FLOPs. Second, the inverse of a $K\times K$ matrix needs to be computed, which can efficiently be done by first computing a Cholesky factorization, followed by taking the inverse of the resulting lower triangular matrix [71]. This requires $\frac{1}{2}K^3+\frac{3}{2}K^2$ complex multiplications, $\frac{1}{2}K^3-\frac{1}{2}K^2$ complex additions and the computation of K square-root operations, which is equivalent to K^3+K^2+K complex FLOPs. Finally, a matrix multiplication of an $M\times K$ with an $K\times K$ matrix needs to be

computed, which requires MK^2 complex multiplications and $MK(K-1)$ complex additions which results in $2MK^2 - MK$ complex FLOPs. This results in a total number of $2MK^2 + \frac{1}{2}K^3 + \frac{3}{2}K^2$ complex multiplications and $2MK^2 + \frac{1}{2}K^3 - \frac{3}{2}K^2$ complex additions. This is equivalent to $8MK^2 + 2K^3 + 6K^2$ real multiplications, $8MK^2 + 2K^3$ real additions or $16MK^2 + 4K^3 + 6K^2$ real FLOPs. This scales as $\mathcal{O}\left(MK^2 + K^3\right)$ FLOPs.

3.4.2 Complexity of Distortion-aware Beamforming Precoder

The DAB precoder executes an iterative optimization procedure for $P = 50$ different initializations. For each initialization, (3.28) is computed for $I = 1000$ iterations. This requires $2MK$ real additions and $2MK$ real multiplications, not taking into account the computation of the gradient $\nabla_{\boldsymbol{W}} R_{\text{sum}}\left(\boldsymbol{W}^{(i)}\right)$. In order to compute the gradient, [44] employs a finite differences approximation which is computed for each element as

$$\begin{aligned}[\nabla_{\boldsymbol{W}} R_{\text{sum}}(\boldsymbol{W})]_{m,k} \approx & \frac{R_{\text{sum}}(\boldsymbol{W} + \boldsymbol{\Delta}_{\Re}^{(m,k)}) - R_{\text{sum}}(\boldsymbol{W})}{\delta} \\ & + j\frac{R_{\text{sum}}(\boldsymbol{W} + \boldsymbol{\Delta}_{\Im}^{(m,k)}) - R_{\text{sum}}(\boldsymbol{W})}{\delta},\end{aligned}$$

where $\boldsymbol{\Delta}_{\Re}^{(m,k)} \in \mathbb{R}^{M \times K}$ is a matrix containing only zeros except for δ at row m and column k, where δ is a small positive constant. Similarly $\boldsymbol{\Delta}_{\Re}^{(m,k)} \in \mathbb{R}^{M \times K}$ contains all zeros except for $j\delta$ at row m and column k. The computation of the additions, subtractions and quotient in this equation requires 4 real additions and 2 real multiplications. Additionally, in this equation the sum rate needs to be evaluated 3 times according equations (3.14), (3.21), (3.17), (3.18) and (3.19). Equation (3.14) requires $K(K-1)$ complex additions, (3.21) requires $(K+1)(M^2-1)$ complex additions, 2 real additions, $(K+1)(M^2-1)+1$ complex multiplications and 1 real multiplication. Equation (3.17) requires $M^2(K-1)+5M$ complex additions and M^2K+9M complex multiplications, (3.18) requires $5M^3-5M^2+4M$ complex additions and $5M^3+15M$ multiplications and (3.19) requires $10M$ additions and $10M$ multiplications. This leads to the following number of real floating-point additions

$$\begin{aligned}& PI[60M^4K^2 + 24M^3K^3 + 12M^3K^2 + 12M^2K^3 \\ & + 420M^2K^2 + 15MK^2 + 4MK],\end{aligned}$$

Table 3.1: Complexity of the GNN, ZF and DAB precoders. Example for the values $P = 50$, $I = 1000$, $M = 64$, $K = 4$, $d_l = 128$ and $L = 8$.

	GNN	GNN (Serial Operations)	DAB	ZF
Additions	$\mathcal{O}\left(MKd_l^2L + M^2Kd_lL + MK^2d_lL\right)$	$\mathcal{O}\left(d_l^2L + Md_lL + Kd_lL\right)$	$\mathcal{O}\left(PIM^4K^2 + PIM^3K^3\right)$	$\mathcal{O}(K^3 + K^2M)$
Multiplications	$\mathcal{O}\left(MKd_l^2L\right)$	$\mathcal{O}\left(d_l^2L\right)$	$\mathcal{O}\left(PIM^4K^2 + PIM^3K^3\right)$	$\mathcal{O}(K^3 + K^2M)$
FLOPs	$\mathcal{O}\left(MKd_l^2L + M^2Kd_lL + MK^2d_lL\right)$	$\mathcal{O}\left(d_l^2L + Md_lL + Kd_lL\right)$	$\mathcal{O}\left(PIM^4K^2 + PIM^3K^3\right)$	$\mathcal{O}(K^3 + K^2M)$
Example (∼FLOPs)	$51{\cdot}10^6$	$20\cdot10^4$	$14{\cdot}10^{12}$	1000

and the following number of real floating point multiplications

$$PI[60M^4K^2 + 24M^3K^3 - 24M^3K^2 + 6M^2K^3$$

$$+ 324M^2K^2 - 3MK^3 + 3MK^2 + 6MK].$$

This scales as $\mathcal{O}\left(PIM^4K^2 + PIM^3K^3\right)$ real floating point additions and $\mathcal{O}\left(PIM^4K^2 + PIM^3K^3\right)$ real floating point multiplications or $\mathcal{O}\big(PIM^4K^2 + PIM^3K^3\big)$ FLOPs.

3.4.3 Numerical Comparison

An overview of the complexity of the GNN and the different benchmarks is given in Table 3.1. Additionally, a numerical example of the number of FLOPs is given for the following case: $P = 50$, $I = 1000$, $M = 64$, $K = 4$, $d_l = 128$ and $L = 8$. This illustrates that the GNN precoder has a complexity that is six orders of magnitude smaller as compared to the DAB precoder. Moreover, when the ability of the GNN to parallelize the computations across the edges is utilized, the number of FLOPs that need to be executed in serial is eight orders of magnitude smaller as compared to the DAB precoder. When accounting for the absolute number of FLOPs, the GNN complexity is significantly higher as compared to ZF ($\sim 50.000\times$). Parallelization of the GNN reduces the number of serial FLOPs to approximately 200 times that of the ZF precoder. However, it has to be noted that this requires additional hardware and as such additional energy consumption.

To keep a concise and fair comparison between the GNN and the benchmarks, only the number of FLOPs is taken into account. However, specialized NN accelerators can improve the inference times of NNs and reduce the energy required to run them [72, 69]. Next to this, a myriad of acceleration techniques can be applied to neural networks, such as tensor decomposition, pruning, general matrix multiplication, etc. [72]. Additionally, NNs are known to be able to work on a lower bit resolution, leading to the use of fixed-point arithmetic rather than floating-point arithmetic, which leads to even faster inference times [72].

Table 3.2: PA parameters of the 11^{th} order polynomial PA model at different back-off values ($\beta_1 = 1$).

IBO [dB]	$\beta_3\,(\cdot 10^{-2})$	$\beta_5\,(\cdot 10^{-3})$	$\beta_7\,(\cdot 10^{-5})$	$\beta_9\,(\cdot 10^{-7})$	$\beta_{11}\,(\cdot 10^{-9})$
-9	-4.38184836 -10.1466832j	1.50490437 8.422084885j	-3.13452827 -28.1868627j	3.49967293 42.06333106j	-1.59432984 -23.1868139j
-7.5	-5.79334438 -9.36769411j	2.39315994 7.94859107j	-5.57663136 -26.92641291j	6.65066314 40.4837957j	-3.14808144 -22.4280442j
-6	-7.50994886 -8.42352484j	3.66782506 7.26453523j	-9.54049052 -24.8371067j	12.2703316 37.5613932j	-6.13183499 -20.8924283j
-4.5	-9.35828409 -7.41305601j	5.16172165 6.46522185j	-14.4481282 -22.2483069j	19.4963213 33.7874265j	-10.0752209 -18.8479147j
-3	-11.1143930 -6.30816977j	6.60156653 5.47141526j	-19.1451680 -18.6610370j	26.2822435 28.0380833j	-13.6811147 -15.4579691j
-1.5	-12.903319 -5.49758824j	8.21176444 4.85204392j	-24.8588087 -16.8144990j	35.2215545 25.6527492j	-18.8139985 -14.3562319j
0	-14.4473655 -4.67375592j	9.58442261 4.13617338j	-29.6362436 -14.3570171j	42.5309097 21.9271142j	-22.9128062 -12.2805850j

Exploration of these avenues can lead to faster and more efficient execution of the GNN.

3.5 Simulation Results

For the following simulations, the polynomial coefficients are obtained by a least squares regression of an 11^{th} order model to the modified Rapp model as described in Section 3.2.2. This regression is performed at the desired IBO, the polynomial coefficients corresponding to the IBO value can be found in Table 3.2. For some specific cases a 3^{th} order PA model at an IBO $= -3$ dB is considered, the PA parameters for this model are[6] $\beta_1 = 1, \beta_3 = -77.82 - 40.12j \cdot 10^{-3}$. For all simulations, the total transmit power is $P_T = M$. Hence, the average power at the input of each PA is $p_{\text{in}} = P_T/M = 1$. The linear PA gain is set to one. Training and testing are done at an IBO of -3 dB, unless specified otherwise, this saturates the PAs significantly more than current cellular systems that require 9-12 dB back-off. After training, the NNs are evaluated based on the sum rate given in (3.14) where the SNIDR is computed numerically according to (3.15). The training set consists of 500 000 generated channels. These are generally

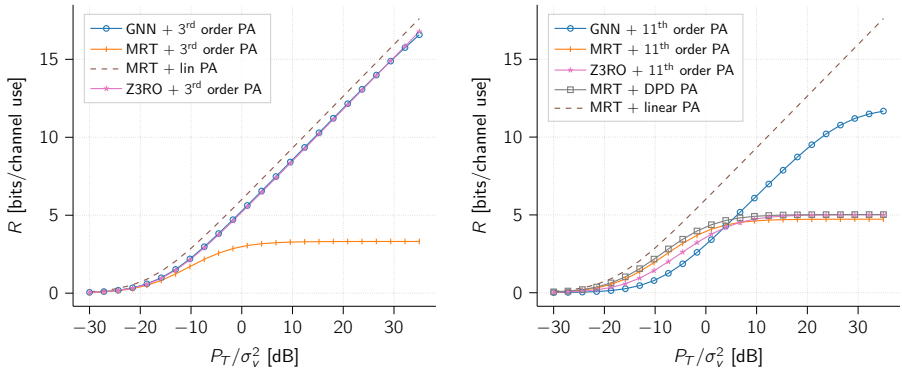
[6] Obtained using a least squares regression of the 3^{th} order PA model to the modified Rapp model at an IBO $= -3$ dB.

sampled from a complex normal distribution with zero mean and variance one $[\boldsymbol{H}]_{i,j} \sim \mathcal{CN}(0,1)$ if a Rayleigh fading channel is assumed. When considering radiation patterns, a deterministic LoS channel is generated according to (3.12), where the user angle (in degrees) is randomly sampled from a discrete uniform distribution $\theta_k \sim \mathcal{U}\{0,180\}$. The hyperparameters of the neural networks are selected by using a validation set of size 2000, while for the simulations performed in this section, an independent test set of size 10 000 is used. For training, a batch size of 64 is used, with an initial learning rate of 5×10^{-3}, which is reduced if the validation loss reaches a plateau. The network is trained for 50 epochs with early stopping if the validation loss does not further decrease. During training, the validation and training loss are monitored to ensure that the neural network has converged to a stable solution and that no overfitting occurs.

3.5.1 Single-User Case

In this section the performance of the GNN as described in Section 3.3.2 is evaluated, for the single-user case $K = 1$. This GNN takes as input the channel matrix $\boldsymbol{H}$ and outputs the precoding matrix $\hat{\boldsymbol{W}}$. For the evaluations performed in this section, the GNN consists of 8 layers (i.e., $L = 8$), with each hidden layer having $d_l = 128$ features. Each GNN-layer is described by (3.35). Where the non-linear activation function in each layer is a leaky rectified linear unit (LReLU), except in the last GNN-layer where a linear activation is used. After this linear activation, the power normalization layer as described in (3.40) is performed. During training, P_T/σ_v^2 is set to 20 dB.

For the single-user case when one assumes the simplified third-order polynomial PA model (i.e., $N = 1$ in (3.7)) the closed-form Z3RO precoder as studied in Chapter 2 can be considered. Therefore, first it is investigated whether the GNN precoder can achieve the same performance as the optimal Z3RO precoder under these assumptions. In Figure 3.6a the achievable rates of the GNN, Z3RO and MRT precoder are depicted as a function of P_T/σ_v^2. The MRT precoder is designed under the assumption of a linear PA. As a consequence, when a non-linear PA is present, the user performance degrades as compared to the linear case, as can be seen in Figure 3.6a. Additionally, it can be seen in Figure 3.6a that the rate of the Z3RO precoder is not limited by distortion but grows logarithmically with P_T/σ_v^2. Indeed, the Z3RO precoder is able to mitigate the third-order distortion in the user direction by saturating one (or a few) antennas with an opposite phase shift, which comes at the cost of a small reduction in array gain. When comparing the GNN precoder, trained on a third-order PA model, with the Z3RO precoder, we see that the GNN achieves similar performance as the Z3RO precoder. Indeed,



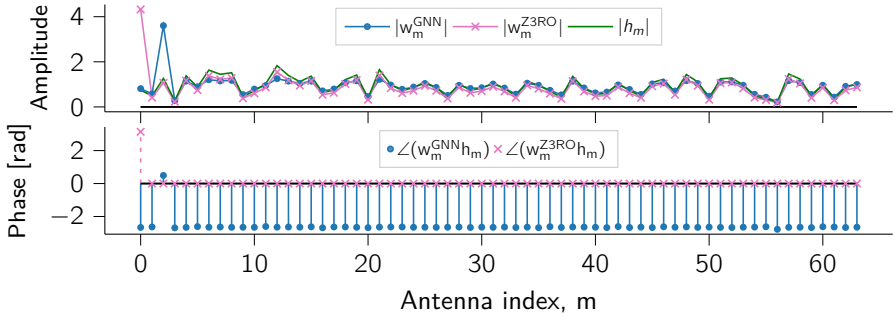
(a) 3rd order PA (b) 11th order PA

Figure 3.6: Achievable rates averaged over the channel realizations taken from the test set evaluated on a 3rd order PA (a) and an 11th order PA (b). Comparing the GNN precoder against MRT, MRT plus a perfect DPD and the Z3RO precoder from [39] for $M = 64$ and $K = 1$. Observations: (a) GNN closely matches the optimal Z3RO precoder. (b) GNN copes better with high-order effects in the PA.

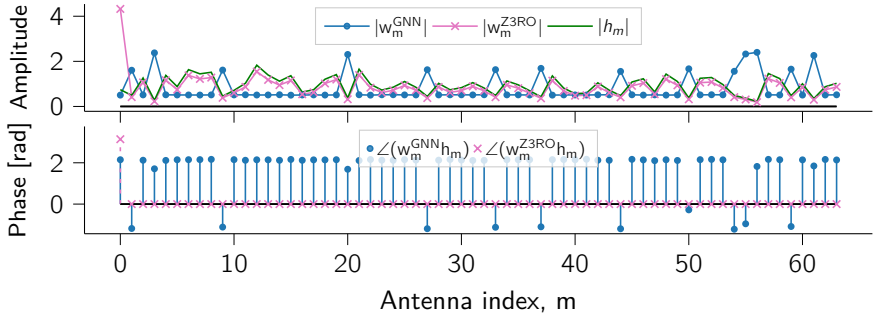
when comparing the amplitude and phase of both precoders in Figure 3.7a, it is clear that the GNN also saturates one or a few of the antennas with an opposite phase shift. In conclusion, for $K = 1$, the GNN has learned a similar precoding structure as the optimal Z3RO precoder.

Second, the case of a high-order PA model is studied, namely an eleventh order PA ($N = 5$). In this case, the Z3RO precoder is no longer optimal as can be seen in Figure 3.6b. Under this PA model, the rate of the Z3RO precoder is limited by distortion as P_T/σ_v^2 grows, this is also the case for the MRT precoder even when it is combined with an ideal DPD. When comparing against the GNN precoder, trained on an eleventh order PA, Figure 3.6b illustrates how the GNN drastically outperforms MRT, Z3RO precoding and even MRT coupled with a perfect DPD. The GNN provides a rate which nearly grows logarithmically with P_T/σ_v^2 and is only limited by distortion for very large P_T/σ_v^2. Note that the GNN precoder only outperforms the benchmarks for large P_T/σ_ν^2 i.e., when the system is mostly corrupted by distortion and not by noise.

When looking at the amplitude and phase of the GNN precoder trained on the 11th order PA in Figure 3.7b it is clear that it no longer behaves as the Z3RO precoder (i.e., perform MRT precoding with one or a few saturated antennas with opposite phase shift to cancel distortion). Rather, it keeps the amplitude of nearly all transmit antennas constant except for a number of antennas that are



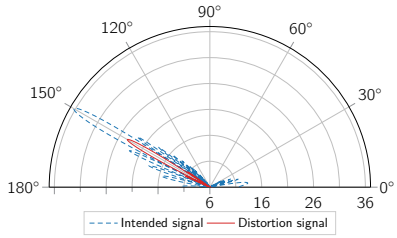
(a) GNN trained on 3rd order PA



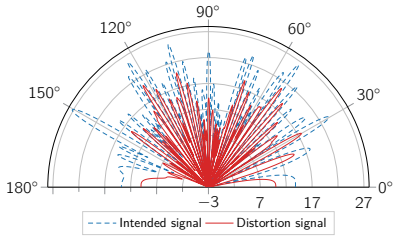
(b) GNN trained on 11thorder PA

Figure 3.7: Amplitude and phase plot per antenna, GNN compared to the Z3RO precoder from [39]. When trained on the third-order model (a), the GNN replicates the behavior of the optimal Z3RO precoder i.e., saturate one antenna with an opposite phase shift to cancel the third order distortion. When trained on an eleventh order PA (b), the GNN saturates more antennas in order to spatially suppress the high-order distortion.

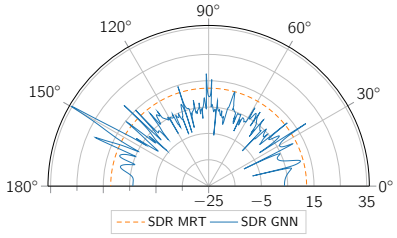
saturated with a different phase shift. This avoids sending all distortion in the user direction as can be seen in the radiation patterns in Figure 3.8. In Figure 3.8a, the radiation patterns of the intended and distortion signals are depicted for the MRT precoder. Here it is clear that the distortion is beamformed in the same direction as the intended signal, i.e., the user direction. Figure 3.8b, depicts the radiation patterns of the intended and distorted signals for the proposed GNN precoder. This figure illustrates that the distortion is more uniformly distributed for the GNN precoder and even close to zero in the user direction. The cost of this dispersed distortion is a reduction in array gain in the user direction. However, as Figure 3.8c illustrates, the SDR radiation pattern for the GNN precoder has a



(a) $P_{\text{lin}}(\tilde{\theta})$, $P_{\text{dist}}(\tilde{\theta})$ for MRT



(b) $P_{\text{lin}}(\tilde{\theta})$, $P_{\text{dist}}(\tilde{\theta})$ for GNN



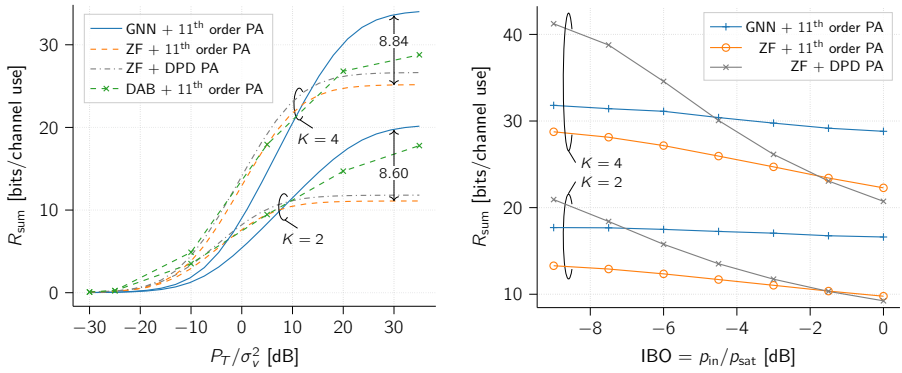
(c) $P_{\text{SDR}}(\tilde{\theta})$ for MRT and GNN

Figure 3.8: Radiation pattern of the intended signal $P_{\text{lin}}(\tilde{\theta})$ and distortion signal $P_{\text{dist}}(\tilde{\theta})$ [dB] for an 11^{th} order PA, $K = 1$ of the GNN precoder (a) and ZF precoder (b) for a pure LoS channel and a half-wavelength ULA. User angle is $\theta = 150°$. In (c) the signal-to-distortion ratio (SDR) radiation pattern $P_{\text{SDR}}(\tilde{\theta})$ is depicted [dB].

clear peak in the user direction, while it is uniform for the MRT precoder.

3.5.2 Multi-User Case

In this section, the same GNN model as in the previous section is used, however, this time the multi-user case is considered. When multiple users are present, both the inter-user interference and distortion to multiple users have to be mitigated. In this more complex scenario, there is no closed-form solution available for the optimal precoder. As such, the GNN is trained to learn how to perform this task. Additionally, the DAB precoder from [44] is considered, as it provides a way to solve the non-convex optimization problem, without any guarantees on optimality. Note that in [44] the optimization procedure is repeated for $P = 50$ different initializations and executed for $I = 100$ iterations per initialization. In [44] this procedure is reported to converge after $I = 100$ iterations. However, given the added complexity of our scenario (i.e., 64 vs. 16 Tx antennas and an 11^{th}-order vs. a 5^{th}-order PA) our simulations did not converge after $I = 100$ iterations.



(a) (b)

Figure 3.9: Achievable rates averaged over the channel realizations taken from the test set evaluated on an 11^{th} order PA. Comparing the GNN precoder against ZF and the DAB precoder from [44] for $M = 64$ and $K \in \{2, 4\}$ at IBO $= -3$ dB in (a). Note that the DAB precoder was only evaluated on 10 channel realizations to reduce simulation time. In (b) the achievable sum rates are plotted for $M = 64$, $K \in \{2, 4\}$, $P_T/\sigma_\nu^2 = 20$ dB with varied IBO. Comparing the GNN, ZF and ZF plus DPD. The GNN is retrained at each IBO point.

Consequently, the simulation results for the DAB precoder in this study are obtained by running the optimization procedure for $I = 1000$ iterations, repeated for $P = 50$ different initializations. In order to compensate for this added simulation time the DAB precoder is only evaluated on 10 channel realizations taken from the test set, while the GNN is evaluated on the full test set[7].

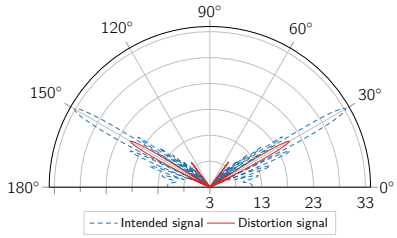
In Figure 3.9a, a comparison between the GNN, DAB, ZF and ZF precoder coupled with a perfect DPD is made. The sum rate is depicted in function of P_T/σ_ν^2 for $K \in \{2, 4\}$, when using the eleventh order PA model. Comparing the GNN precoder against the ZF precoder, for $K = 2$ the sum rate is increased by 8.60 bits/channel use at $P_T/\sigma_\nu^2 = 30$ dB, when using the proposed method, and 8.84 bits/channel use for $K = 4$. The GNN precoder shows superior performance over the DAB precoder. Next to this, it is shown that the proposed solution outperforms a perfect DPD combined with ZF, when the system is highly distortion limited (i.e., for high P_T/σ_ν^2 levels). This shows the ability of the GNN to cancel high-order non-linear distortion in the multi-user scenario, which results in significant

[7]The results when evaluating the GNN on the same 10 channel realizations as the ones used to evaluate the DAB precoder match the results obtained when evaluating the GNN on the full test set.

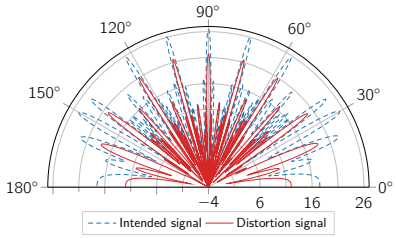
increases in capacity. Additionally, this illustrates that the higher the number of users becomes, the less gain is to be obtained, in terms of added rate per user, by using the GNN precoder. This is expected as non-linear distortion is more spatially spread out when more users are present [34]. In other words, less distortion is beamformed in the user directions, which leads to less potential gains for mitigating this distortion. Moreover, when more users are present, canceling all distortion to all users becomes more complex. Nevertheless, the GNN-based precoder achieves a significant increase in achievable sum rate as compared to the benchmark approaches.

Figure 3.9b depicts the sum rate when p_{in} is fixed but p_{sat} is varied, resulting in a varied IBO. This shows that, for $K \in \{2, 4\}$, the GNN always outperforms the classical ZF precoder, when evaluated using the eleventh order PA model. Moreover, when $K = 2$, the GNN is able to achieve a nearly constant sum rate over a wide range of IBO. This illustrates the ability of the GNN to suppress nearly all high-order distortion in the user directions. When comparing the GNN to ZF plus a perfect DPD in Figure 3.9b, it is evident that the GNN is most beneficial when a lot of distortion is present, i.e., at low back-off. However, we stress the fact that the GNN does not have to be used as a replacement, but could be used in combination with DPD. When (perfect) DPD is available, the PA characteristic after applying DPD can be modeled as a polynomial on which the GNN can be retrained. The combination of both approaches could produce even better sum rates. However, this would add a significant complexity burden.

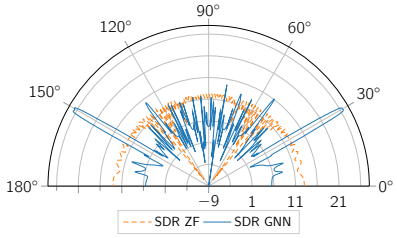
Finally, in Figures 3.10 and 3.11 the radiation patterns for $K = 2, 4$ are depicted. In Figures 3.10a and 3.10b, the radiation patterns of the intended and distorted signals are plotted for the ZF and GNN precoder respectively, for $K = 2$. Figure 3.10a clearly illustrates how for ZF precoding the distortion is beamformed in both user directions. Figure 3.10b reveals that for GNN precoding the distortion is mainly beamformed in the non-user directions, as the distortion in the user directions is very small. This however comes at the cost of a reduction in array gain for the linearly amplified signal. Nonetheless, in Figure 3.10c the SDR radiation pattern is presented, which shows two clear peaks in the user directions, for the GNN precoder, while the SDR for the ZF precoder is nearly uniformly distributed. In Figures 3.11a and 3.11b, the radiation patterns of the intended and distorted signals are plotted for the ZF and GNN precoder respectively, for $K = 4$ users. For the ZF precoder, the distortion is again beamformed in the user directions while it is more spatially spread out for the GNN precoder. When looking at the SDR radiation pattern in Figure 3.11c, we see a slight improvement of the SDR in the user directions for the GNN precoder as compared to the ZF precoder.



(a) $P_{\text{lin}}(\tilde{\theta})$, $P_{\text{dist}}(\tilde{\theta})$ for ZF



(b) $P_{\text{lin}}(\tilde{\theta})$, $P_{\text{dist}}(\tilde{\theta})$ for GNN



(c) $P_{\text{SDR}}(\tilde{\theta})$ for ZF and GNN

Figure 3.10: Radiation pattern of the intended signal $P_{\text{lin}}(\tilde{\theta})$ and distortion signal $P_{\text{dist}}(\tilde{\theta})$ [dB] for an 11^{th} order PA, $K = 2$ of the GNN precoder (a) and ZF precoder (b) for a pure LoS channel and a half-wavelength ULA. User angles are $\theta \in \{30°, 150°\}$. In (c) the signal-to-distortion ratio (SDR) radiation pattern $P_{\text{SDR}}(\tilde{\theta})$ is depicted [dB].

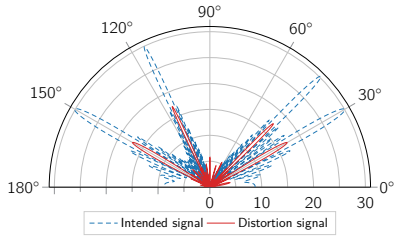
3.5.3 Improvements in Power Consumption

PA Power Consumption

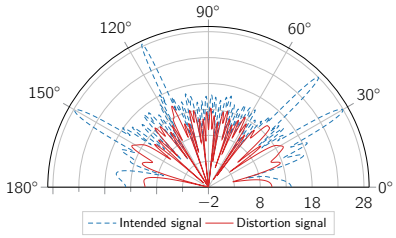
By spatially suppressing the non-linear distortion, a lower IBO can be used to achieve a desired sum rate. By allowing the PAs to work closer to saturation, their energy efficiency is improved. In this section the improvements in PA power consumption are quantified. For a class B amplifier the following model of the energy efficiency can be used [73]. The efficiency of PA m is defined as

$$\eta_m = \eta_{\max} \sqrt{\frac{p_m}{p_{\text{sat}}}}. \tag{3.45}$$

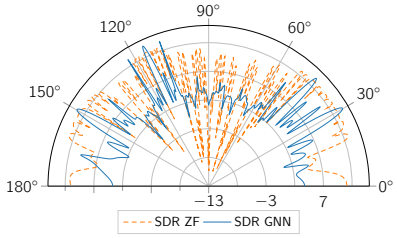
Here p_{sat} is the saturation power of the PA, while $\eta_{\max}$ is the maximum achievable efficiency, for a class B amplifier $\eta_{\max} = 0.785$ [73]. Next to this, p_m is defined as



(a) $P_{\text{lin}}(\tilde{\theta})$, $P_{\text{dist}}(\tilde{\theta})$ for ZF



(b) $P_{\text{lin}}(\tilde{\theta})$, $P_{\text{dist}}(\tilde{\theta})$ for GNN



(c) $P_{\text{SDR}}(\tilde{\theta})$ for ZF and GNN

Figure 3.11: Radiation pattern of the intended and distortion signal $P_{\text{lin}}(\tilde{\theta})$, $P_{\text{dist}}(\tilde{\theta})$ [dB] for an 11^{th} order PA, $K = 4$ of the GNN (a) and ZF precoder (b) for a pure LoS channel and a half-wavelength ULA. User angles are $\theta \in \{30°, 45°, 115°, 150°\}$. In (c) the signal-to-distortion ratio (SDR) radiation pattern $P_{\text{SDR}}(\tilde{\theta})$ is depicted [dB].

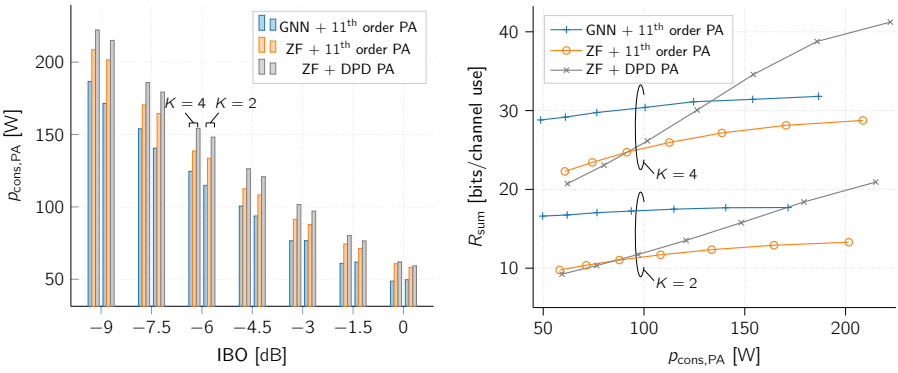
the power at the output of amplifier m

$$p_m = \mathbb{E}\left(|y_m|^2\right) = \mathbb{E}\left(|\phi(x_m)|^2\right) = \mathbb{E}\left(\left|\phi\left(\sum_{k=0}^{K-1} w_{m,k} s_k\right)\right|^2\right). \tag{3.46}$$

The total consumed PA power is then given by

$$p_{\text{cons,PA}} = \sum_{m=0}^{M-1} \frac{p_m}{\eta_m} = \frac{\sqrt{p_{\text{sat}}}}{\eta_{\max}} \sum_{m=0}^{M-1} \sqrt{\mathbb{E}\left(|y_m|^2\right)}. \tag{3.47}$$

From this expression, it can be seen that the consumed power is not proportional to the average input back-off but rather to the sum of the square root of the output power of the PAs. Note that $\mathbb{E}\left(|x_m|^2\right)$ and $\mathbb{E}\left(|y_m|^2\right)$ represent dimensionless digital powers. These can be translated to an actual power value by scaling it with a physical reference power (in watt) as $p_{\text{ref},m}\mathbb{E}\left(|x_m|^2\right)$. Note that in this thesis $p_{\text{ref},m} = 1$ W is assumed. This is inline with sub-6 GHz BSs such as the one in [74], where the power per antenna ranges from 1.25 to 3.125 watts, depending on the exact frequency band.



(a) (b)

Figure 3.12: (a) Consumed PA power in function of IBO (b) achievable sum rate in function of consumed PA power. Both (a) and (b) compare the GNN, ZF and ZF + DPD precoders for $M = 64$, $K \in \{2, 4\}$, $\frac{P_T}{\sigma_\nu^2} = 20$ dB.

In Figure 3.12a the consumed power of the PAs in function of the input back-off is depicted for $K = 2$ and $K = 4$. This figure illustrates that the GNN has a lower PA consumed power as compared to ZF and ZF + DPD precoding, for a same average input back-off (and thus a same transmit power). Intuitively, this lower consumption is due to the fact that the GNN saturates a number of antennas. These saturated antennas have a very high efficiency, while they also receive more power as compared to the less efficient antennas. This reduced power consumption is formally explained by Proposition 1 in Chapter 5.

Moreover, for a fair comparison one needs to select a quality of service constraint, e.g., a sum rate constraint, and compare the consumed power for the different precoders meeting this rate constraint. As an example for $M = 64$, $K = 4$ let us consider a desired sum rate of 30 bits/channel use. From Figure 3.12b the consumed power to achieve this rate can be seen, for GNN precoding this is $p_{\text{cons,PA}} = 85.57\,\text{W}$, for ZF precoding $p_{\text{cons,PA}} = 284.29\,\text{W}$ and for ZF precoding + DPD this is $p_{\text{cons,PA}} = 125.98\,\text{W}$.[8] This reveals that the GNN precoder has a PA power consumption that is 3.32 times smaller as compared to ZF precoding and 1.47 times smaller as compared to ZF precoding plus DPD.

[8]Obtained using linear interpolation between the points of Figure 3.12b.

Digital Signal Processing Power Consumption

In this section the consumed power of the hardware accelerator needed to run the GNN is quantified. Considering the scenario where $M = 64$, $K = 4$, based on Section 3.4 it is clear that 164 MFLOPs are required to compute one forward pass of the GNN. In order to establish the desired speed (FLOPs/s) of the hardware accelerator, the coherence time of the channel is computed to determine the time available to compute the precoding coefficients. The coherence time is proportional to the inverse of the maximum Doppler frequency $f_m = \frac{v}{c} f_c$ and is often defined as $T_c = \frac{1}{2f_m}$ [19]. Assuming a carrier frequency of $f_c = 5$ GHz and a receiver velocity of $v = 10$ m/s, this gives a coherence time of 3 ms. As a design choice, it is assumed that 10% of the coherence time can be used to compute the precoding coefficients. For instance, in [75, 76], a latency margin of 153 µs is available for precoding within a 0.5 ms radio slot (supporting a minimum coherence time of 430 µs), equating to 35% of the coherence time that is available for precoding. Given our larger coherence time of 3 ms, a conservative 10% is adopted. This establishes a required hardware accelerator speed of 549 giga floating point operations per second (GFLOPs/s). Based on [69] the state-of-the-art Manticore accelerator [77] is selected, which can achieve the desired speed and precision.[9] The full Manticore design consists out of 4096 cores, however, in [77] a prototype with 24 cores was developed and measured. This prototype is able to achieve 50 GFLOPs/s with a power consumption of 1 W. Given the parallel architecture of the GNN it can be deployed across multiple cores. Hence, in order to reach the desired speed of 549 GFLOPs/s 264 cores can be utilized rather than 24. This leads to a total digital signal processing (DSP) power consumption of 10.98 W.

Total Power Consumption

An overview of the total consumed power (PA and processing power) for the GNN can be seen in Table 3.3. It can be seen that the total consumed power for GNN precoding is 96.55 W, for ZF precoding 284.29 W and for ZF + DPD 125.98 W. Note that, as a worst-case scenario, the processing power of ZF and DPD is assumed to be zero, while in reality this processing power will be substantial when DPD is implemented. Even in this scenario, the GNN precoder consumes 2.95 times less power as compared to ZF and 1.31 times less power as compared to ZF+DPD.

[9] Note that the GNN considered in this work requires single precision (32 bit) floating point operations.

3.5.4 Validation on Rapp PA Model

In this section, the GNN precoder which is trained on the eleventh order PA model is evaluated on the modified Rapp model. The modified Rapp model is originally used to obtain the coefficients of the polynomial model by performing a least squares regression, as discussed in Section 3.2.2. By evaluating the GNN on the modified Rapp model, the effects of using a polynomial model, rather than the actual PA characteristic can be studied. In Figure 3.13 the rate is depicted in function of P_T/σ_ν^2 when the modified Rapp model is used for the amplification stage. For the single-user case $K = 1$, this figure illustrates how the GNN precoder trained on a polynomial PA still outperforms the MRT and Z3RO precoders when the underlying modified Rapp model is assumed. For the multi-user case $K \in \{2, 4\}$ in Figure 3.13, it is again clear that the GNN precoder outperforms the ZF precoder when the modified Rapp model is deployed as the amplification model. This highlights the fact that the polynomial model can successfully be used to train the GNN even if the underlying PA characteristic is slightly different.

3.5.5 Leveraging SNR Regime Information

In this section the performance of the GNN as described in Section 3.3.3 is evaluated. This GNN takes as input edge features the channel coefficients and the normalized SNR, and is denoted as GNN_{SNR}. GNN_{SNR} is compared against GNN_{base} which only takes the channel matrix as input. The GNN architecture is kept the same as in Section 3.3.2 except for the case where $K = 4$ where $d_l = 256$ features per layer are used rather than 128.

In Figure 3.14a and 3.14b the sum rate is depicted in function of P_T/σ_ν^2 for $K \in \{1, 2, 4\}$, for the eleventh order PA model. This figure illustrates the ability of GNN_{SNR} to leverage the extra SNR input to deliver performance over the full range of P_T/σ_ν^2. This alleviates the need for retraining the GNN at each value of

Table 3.3: Consumed power of the GNN, ZF and ZF + DPD precoders for $M = 64$, $K = 4$, for a fixed data rate of 30 bits/channel use. As a worst-case scenario the processing power for ZF and DPD are assumed to be zero.

Precoder	Precision	Accelerator power [W]	$p_{PA,cons}$ [W]	total power [W]
GNN + Manticore [77]	single precision (SP)	10.98	85.57	96.55
ZF + DPD	*NA*	*NA*	125.98	125.98
ZF	*NA*	*NA*	284.29	284.29

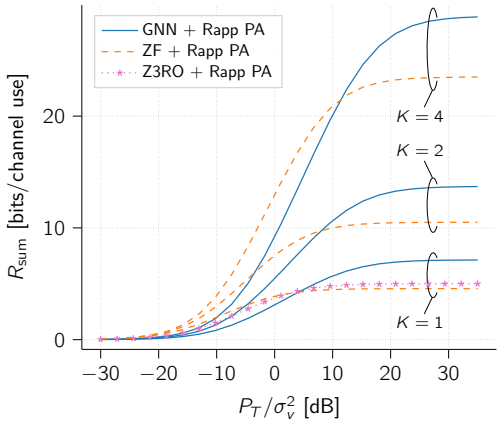
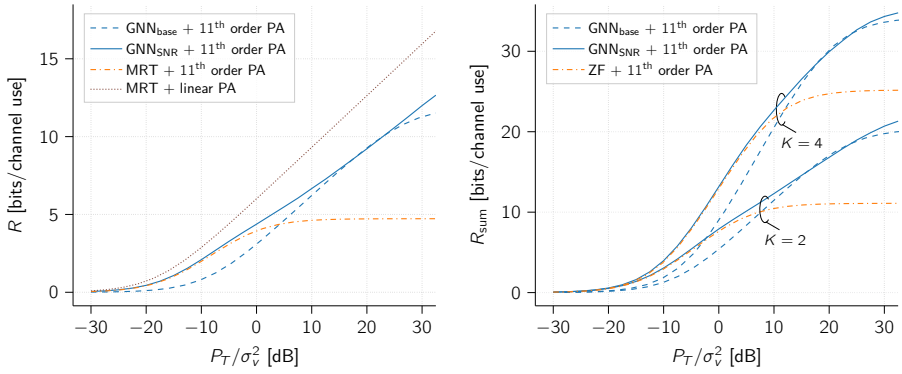


Figure 3.13: Achievable sum rates averaged over 1000 Rayleigh channel realizations taken from the test set evaluated using the Rapp PA model (3.9, 3.10). $M = 64$, $K \in \{1, 2, 4\}$, with a varied P_T/σ_ν^2.



(a) $K = 1$ (b) $K \in \{2, 4\}$

Figure 3.14: Achievable rates for $M = 64$, $K \in \{1, 2, 4\}$ and an 11^{th}order PA. Comparing the base GNN from Section 3.3.2 against the GNN that leverages SNR information from Section 3.3.3.

P_T/σ_ν^2. In this way, GNN_{SNR} not only performs well in the distortion limited regime (i.e., for high P_T/σ_ν^2) but is also capable of delivering the same performance as ZF in the noise limited regime (i.e., for low P_T/σ_ν^2), which GNN_{base} was not able to do.

3.6 Conclusion

In this chapter, a massive MIMO system is considered, where a GNN is trained to learn a mapping between the channel and precoding matrix which maximizes the sum rate in the presence of high-order non-linear PA distortion. The complexity of the proposed solution is analyzed and shown to be six orders of magnitude lower as compared to the benchmark DAB precoder. Simulation results indicate that the proposed solution offers an increased capacity as compared to ZF in the distortion-limited regime, i.e., an energy-efficient regime. For instance, at an IBO of -3 dB the proposed precoder produces an increase in sum rate of 8.60 and 8.84 bits/channel use for the two and four user cases respectively. Additionally, it is shown that the proposed solution achieves a higher sum rate as compared to ZF precoding plus DPD, as DPD is limited by the distortion caused by saturation based clipping. Radiation patterns demonstrate that the higher rate is achieved by spatially suppressing the distortion in the user-directions. By utilizing the proposed precoder, PAs can operate closer to saturation, which increases their energy efficiency. For instance, it is shown that, in the four user-case, for a fixed sum rate the PA consumed power of the proposed precoder is 3.32 times and 1.47 times lower as compared to ZF precoding and ZF plus DPD respectively. In addition, it is shown that the PA power consumption reduction is achieved at the price of a small increase in DSP consumption so that the total power consumption is still greatly reduced. Furthermore, the GNN is extended to incorporate additional knowledge, i.e., the SNR regime, which alleviates the need to retrain at each SNR point. These results are especially promising in the multi-user case where no closed-form/low-complexity solution to the distortion-aware precoding problem is available. At the same time, understanding how these GNN-based precoders behave under varying channel conditions and extending them to frequency-selective channels and OFDM systems introduces additional challenges. These aspects are explored in the next chapters. In Chapter 4, the impact of varying channel statistics on GNN-based precoding is studied, while in Chapter 5 an extension for OFDM systems is made, in order to handle frequency selectivity of the channel.

Chapter 4

Generalization to Varying Channel Conditions

In the previous chapter, a graph neural network (GNN) was proposed to perform distortion-aware precoding. This chapter examines the effect of domain/covariate shift on neural network (NN)-based precoding. Covariate shift occurs when the data distribution used during training (training distribution) is different from the data distribution observed during the use phase of the model (test distribution). When the model fails to generalize to the (different) unseen test distribution, this phenomenon results in a performance drop. This chapter studies whether such a performance drop occurs and how severe it is. This aligns with research question (Q2.2) What is the impact of varying channel conditions (different training distribution vs. testing distribution) on the performance of NN-based precoders?

4.1 Introduction

A common assumption in machine learning (ML) is that the data that is used to train the model is sampled from the same distribution as the data that the model will experience during inference, i.e., the train and test distributions are the same. In the previous chapter, this assumption was made as the NNs are trained and evaluated on one fixed channel model. Note that even though the same distribution was used during training and testing, the training and testing datasets were strictly separated. However, real-world channel realizations might

have different characteristics. This can be due to model mismatch between the channel model and the real-world channels or due to varying channel conditions. The problem of having a different train and test distribution is known as domain shift or covariate shift [78, 79, 80]. With respect to NN-based precoding, in [81], meta-learning domain generalization [80] is used to improve generalization to new unseen channel environments. In [81], it is reported that a NN-based precoder experiences a performance loss of 53% when trained at one specific base station (BS) site and deployed at another BS site. By leveraging meta-learning domain generalization, this loss is reduced to below 1% at high signal-to-noise ratio (SNR) and 20% at low SNR. Additionally, it is shown that using a limited finetuning dataset (consisting of samples from the test distribution), allows the NN to outperform classical techniques such as the weighted minimum mean squared error (WMMSE) algorithm. Likewise, in our follow-up work [82, 83], a GNN-based precoder was trained on simulated channels. To generalize to the unseen real-life channels, a limited fine-tuning set of measured channels was captured using the techtile testbed [84]. This transfer learning strategy allows for good performance on the real-life channel distribution, while requiring minimal samples from the test distribution. To study the effect of covariate shift resulting from changing channel distributions, this chapter explores what happens when different channel models are considered during both training and testing. For this, the GNN from Chapter 3, is trained on an independently and identically distributed (i.i.d.) Rayleigh fading channel and tested on different channel distributions, namely line-of-sight (LoS) and Rician channel models. Although this chapter focuses on model-based channel distributions, it should be noted that actual real-world channels may still differ from these models. Generalization to unseen real-world channel measurements was studied in our follow-up works [82, 83].

4.2 Problem Formulation

To study the problem of covariate shift in isolation, a simplified system model is proposed. NN-based precoding is considered for the classical massive multiple input multiple output (MIMO) downlink system without hardware impairments, i.e., no non-linear power amplifiers (PAs) are considered in this chapter. The sum rate is considered as a performance metric to assess the performance drop associated with covariate shift.

4.2.1 System Model

A downlink BS is considered that utilizes M transmit antennas to serve K single-antenna users. The symbol for user k is zero-mean circularly symmetric complex Gaussian with unit variance, and is denoted as s_k. The precoded symbol vector is denoted as

$$\boldsymbol{x} = \boldsymbol{W}\boldsymbol{s} \tag{4.1}$$

where $\boldsymbol{W} \in \mathbb{C}^{M \times K}$ is the precoding matrix and the vector $\boldsymbol{s} \in \mathbb{C}^{K \times 1}$ combines the symbols intended for all users. The received signal vector is given by

$$\boldsymbol{r} = \boldsymbol{H}^{\mathsf{T}}\boldsymbol{x} + \boldsymbol{v} \tag{4.2}$$

where $\boldsymbol{H} \in \mathbb{C}^{M \times K}$ is the channel matrix and $\boldsymbol{v} \in \mathbb{C}^{K \times 1}$ contains i.i.d. zero mean complex Gaussian noise samples with variance σ_v^2. The SNR for user k can be expressed as

$$\mathrm{SNR}_k = \frac{|\boldsymbol{h}_k^{\mathsf{T}}\boldsymbol{w}_k|^2}{\sum_{k' \neq k} |\boldsymbol{h}_k^{\mathsf{T}}\boldsymbol{w}_k'|^2| + \sigma_v^2}. \tag{4.3}$$

Next, an achievable sum rate can be computed as

$$R_{\mathrm{sum}} = \sum_{k=0}^{K-1} \log_2(1 + \mathrm{SNR}_k). \tag{4.4}$$

4.2.2 GNN-based Precoder

A GNN-based precoder is considered, which learns the mapping from the channel matrix $\boldsymbol{H}$ to the linear precoding matrix $\boldsymbol{W}$. The learned function is denoted as $\boldsymbol{W} = f(\boldsymbol{H}; \boldsymbol{\theta})$ and is parameterized by $\boldsymbol{\theta}$. The GNN considered in this study has the same structure as in Chapter 3 and is described by equation (3.35). The model is trained in a self-supervised manner by maximizing the expected sum rate, denoted in (4.4), subject to a transmit power constraint $\mathrm{tr}\left(\boldsymbol{W}\boldsymbol{W}^H\right) \leq P_T$. Note that the power constraint is achieved by applying a scalar normalization to the output of the GNN as $\boldsymbol{W}^{norm} = \alpha\boldsymbol{W}$, with $\alpha = \sqrt{P_T / \mathrm{tr}\left(\boldsymbol{W}\boldsymbol{W}^H\right)}$. The GNN considered in this chapter has $L = 6$ layers, meaning it consists of an input layer, 4 hidden layers, and 1 output layer. The number of in- and output features is fixed to $d_0 = d_{L-1} = 2$ and the number of features in each hidden layer is constant at $d_l = 64\ \forall l \in \{2, \cdots, L-2\}$. Each model is trained at an SNR $= P_T/\sigma_v^2 = 20$ dB, for 50 epochs, with a learning rate of 5×10^{-4} and a batch size of 32. The

training set consists of 100.000 channel realizations, the validation set of 1.000 channel realizations and the test set of 10.000 realizations of the selected channel distribution.

4.2.3 Channel models

In this section, we outline the channel models considered for studying covariate shift. Note that large-scale fading is not considered in this work, as it is assumed that different users have the same large-scale fading coefficient. First, a Rayleigh fading channel model is specified as

$$[\boldsymbol{H}]_{m,k} \sim \mathcal{CN}(0,1). \tag{4.5}$$

where $[\boldsymbol{H}]_{m,k}$ represents the wireless channel between base station antenna m and user k.

As a second model, a LoS channel model is considered. This model is defined as

$$[\boldsymbol{H}]_{m,k} = e^{-jm\frac{2\pi}{\lambda_c}d\cos(\theta_k)} \tag{4.6}$$

$$\theta_k \sim \mathcal{U}(0,180). \tag{4.7}$$

where $m\frac{2\pi}{\lambda_c}d\cos(\theta_k)$ models the antenna dependent phase shift when considering a uniform linear array (ULA) and a narrowband system. Next to this, λ_c represents the carrier wavelength, d the antenna spacing, and θ_k the user angle which is sampled from a uniform distribution bounded between 0 and 180 degrees.

Finally, a Rician channel model is considered, which reflects a scenario where a strong dominant path (often corresponding to a LoS component) is present along with multiple scattered paths. This channel model is defined as

$$[\boldsymbol{H}]_{m,k} = \sqrt{\frac{K_{rice}}{K_{rice}+1}}e^{-jm\frac{2\pi}{\lambda_c}d\cos(\theta_k)} + \sqrt{\frac{1}{K_{rice}+1}}h_{\text{NLOS}} \tag{4.8}$$

where $h_{\text{NLOS}} \sim \mathcal{CN}(0,1)$ represents the received non-line-of-sight (NLoS) paths and $\theta_k \sim \mathcal{U}(0,180)$ the user angle of the LoS path. Additionally, K_{rice} represents the ratio between the power in the LoS and NLoS paths. Note that for all channel models, the power of the channel coefficients is normalized to be unit power. An overview of the different channel distributions can be seen in Figure 4.1.

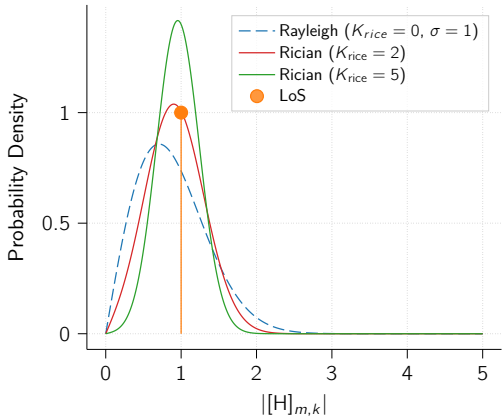


Figure 4.1: Magnitude of the different channel distributions.

4.3 Simulation Results

4.3.1 Methodology

The GNN is trained on one of the three channel models and evaluated on the others, i.e., a different training and test distribution is used, to explore the impact of covariate shift. To indicate which training distribution is used, the training distribution is written as a subscript, hence the different models are referred to as: $\text{GNN}_{\text{Rayleigh}}$, GNN_{LoS} and GNN_{Rice}. Additionally, a model is considered that is trained on an equal mixture of all three channel models, referred to as $\text{GNN}_{\text{mixed}}$. Next to the NN-based precoders, zero forcing (ZF) precoding is considered as a benchmark. For all simulations, the number of transmit antennas is $M = 16$, two users are considered $K = 2$ and the power constraint is $P_T = M$. While $M = 16$ is modest compared to typical large-scale massive MIMO deployments, it captures the key spatial multiplexing and beamforming effects while keeping the numerical simulations computationally tractable.

4.3.2 Rayleigh Training Distribution

When training on a Rayleigh distribution, the NN is exposed to a wide range of input channel magnitudes. Hence, we hypothesize that a wide range of the input space is covered by training on Rayleigh channels, and the influence of covariate shift might be limited. To investigate this, a GNN is trained on i.i.d. Rayleigh

fading channels and tested on LoS and Rician channels. The result of this model is compared to GNNs that are solely trained on Rician and LoS channels. The models that are trained under the same train and test distribution act as an upper bound on the performance. In this way, the model that experiences covariate shift can be compared to these upper bounds.

In Figure 4.2 LoS channels are considered for the test distribution, while $\text{GNN}_{\text{Rayleigh}}$ is trained on a set of Rayleigh fading channels. As a benchmark, ZF is considered. Next to this, a model fully trained on LoS channels (i.e., the test distribution) GNN_{LoS} is added. From this figure, a small performance drop can be observed at high SNR when there is covariate shift from Rayleigh to LoS channels. However, this effect is limited as only 1.5 bits/channel use is lost at an SNR = 20 dB as compared to ZF precoding and the GNN trained on the LoS distribution.

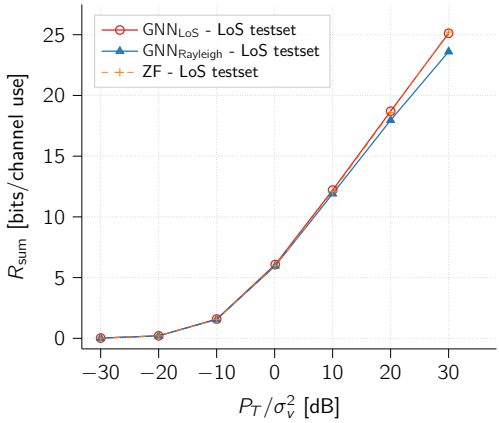


Figure 4.2: $\text{GNN}_{\text{Rayleigh}}$ is trained on Rayleigh fading channels, while GNN_{LoS} is trained on LoS channels. Both models are evaluated on a test set of LoS channels.

In Figure 4.3 $\text{GNN}_{\text{Rayleigh}}$ is trained on a set of Rayleigh fading channels while it is tested on Rician channels. Here we see a minimal performance drop at high SNR when there is covariate shift from Rayleigh to Rician channels. This performance loss is negligible as less than one bit/channel use is lost as compared to ZF. Additionally, it can be noted that even the model trained on the Rician distribution (GNN_{Rice}) experiences a similar performance gap with respect to ZF precoding.

In Figure 4.4, again $\text{GNN}_{\text{Rayleigh}}$ is considered, however this time, a 90% confidence is added to the plot. Figure 4.4a considers LoS channels for the test distribution, while Figure 4.4b considers Rician channels as the test distribution. From these plots, it can be noted that even though the average performance of $\text{GNN}_{\text{Rayleigh}}$ is

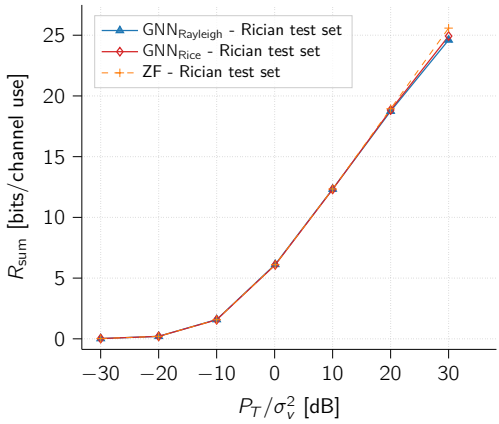


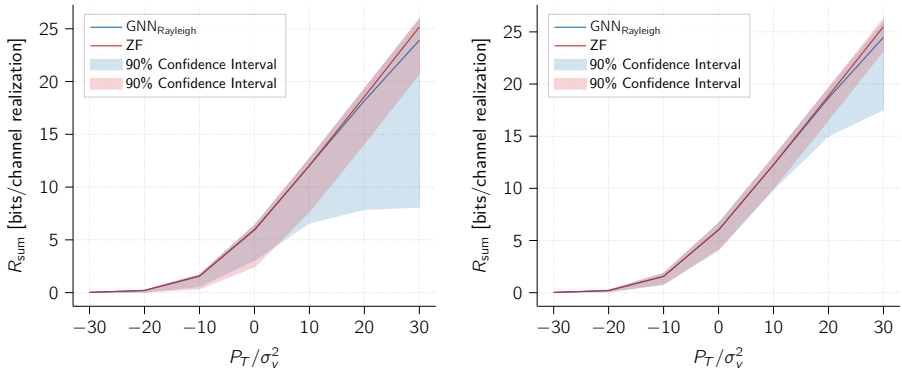
Figure 4.3: GNN$_{\text{Rayleigh}}$ is trained on Rayleigh fading channels, while GNN$_{\text{Rice}}$ is trained on LoS channels. Both models are evaluated on a test set of Rician channels with a rice factor of $K_{rice} = 5$.

only slightly reduced as compared to ZF, the variance in sum rate of GNN$_{\text{Rayleigh}}$ is much larger than that of ZF precoding. From this, it can be noted that covariate shift might have a limited impact on the average performance; however, the worst-case performance on an individual channel realization level is much more significant. This phenomenon can also be observed in Figure 4.6a, where the difference in sum rate between ZF precoding (which is close to optimal when no non-linear PAs are present) and GNN$_{\text{Rayleigh}}$ is considered both on a LoS and Rician test set.

In Figure 4.5 the difference in sum rate between GNN$_{\text{LoS}}$ and GNN$_{\text{Rayleigh}}$ is considered when evaluating on a LoS test set and the difference in sum rate between GNN$_{\text{Rice}}$ and GNN$_{\text{Rayleigh}}$ while testing on Rician channels is considered. Also, here we see a small effect of covariate shift. The models trained on the respective test distributions slightly outperform the model trained only on the Rayleigh distribution. However, the difference in sum rate remains rather small at 0.4 and 1.2 bits/channel use for GNN$_{\text{LoS}}$ and GNN$_{\text{Rice}}$, respectively.

4.3.3 Mixed Training Distributions

To limit the effect of covariate shift, a model can be trained on a mixture of all channel distributions. In Figure 4.6, the difference in sum rate between GNN$_{\text{mixed}}$, a model trained on a mixture of all three channel distributions, and GNN$_{\text{Rayleigh}}$



(a) LoS test set (b) Rician test set

Figure 4.4: GNN$_{\text{Rayleigh}}$ is trained on Rayleigh fading channels. The model is evaluated against ZF on a test set of LoS (a) and Rician channels (b) with a rice factor of $K_{rice} = 5$. $M = 16$, $K = 2$ - GNN: 4 layers, 64 features.

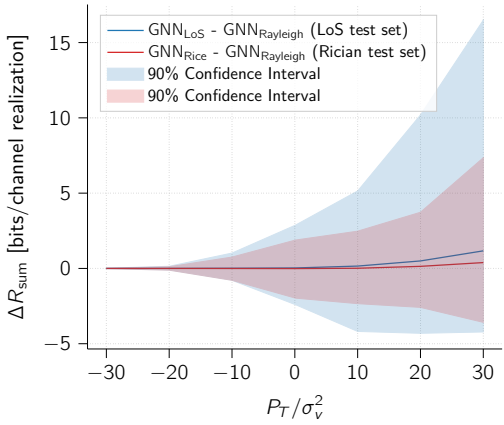
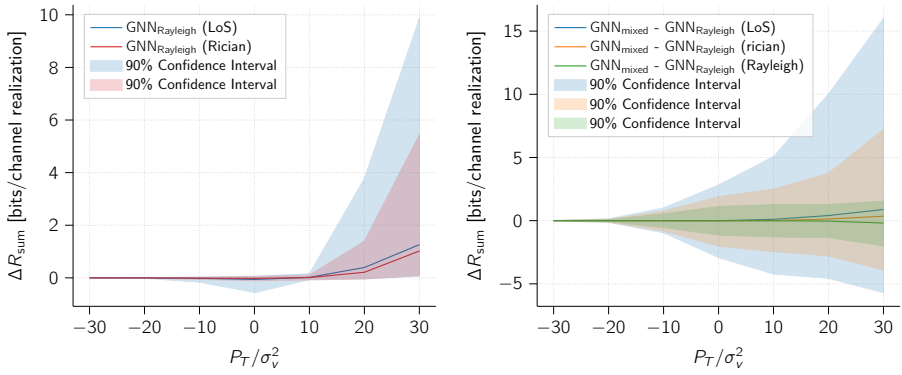


Figure 4.5: .
GNN$_{\text{Rayleigh}}$, GNN$_{\text{Rician}}$ and GNN$_{\text{LoS}}$ are trained on Rayleigh, Rician and LoS channels respectively. When a LoS testset is considered $\Delta R_{\text{sum}} = R_{\text{sum,GNN}_{\text{LoS}}} - R_{\text{sum,GNN}_{\text{Rayleigh}}}$ is computed, when a Rician test set is considered $\Delta R_{\text{sum}} = R_{\text{sum,GNN}_{\text{Rician}}} - R_{\text{sum,GNN}_{\text{Rayleigh}}}$ is computed.

is evaluated on three different test sets, from the three channel distributions. This figure highlights that GNN$_{\text{mixed}}$ outperforms GNN$_{\text{Rayleigh}}$ on both the LoS and Rician test sets. As expected, on the Rayleigh test set GNN$_{\text{Rayleigh}}$ performs



(a) $\Delta R_{\text{sum}} = R_{\text{sum,ZF}} - R_{\text{sum,GNN}_{\text{Rayleigh}}}$ (b) $\Delta R_{\text{sum}} = R_{\text{sum,GNN}_{\text{mixed}}} - R_{\text{sum,GNN}_{\text{Rayleigh}}}$

Figure 4.6: The performance degradation is expressed as $\Delta R_{\text{sum}} = R_{\text{sum,benchmark}} - R_{\text{sum,GNN}_{\text{Rayleigh}}}$. In (a) the considered benchmark is ZF, tested on Rician and LoS channels. In (b) the considered benchmark is $\text{GNN}_{\text{mixed}}$ tested on Rayleigh, Rician and LoS channels. The considered test set is indicated between brackets in the legend.

slightly better as compared to $\text{GNN}_{\text{mixed}}$. This shows that a model trained on a wide range of channel distributions can act as a trade-off in order to minimize the effects of covariate shift.

4.4 Conclusion

In this chapter, the effect of covariate shift, i.e., a difference between the training and testing distribution, is evaluated. Given that the Rayleigh distribution covers a large part of the potential input space of the GNN-based precoder, a model trained on this distribution generalizes well to other channel distributions. On the average performance, we see a limited effect of covariate shift between Rayleigh, Rician, and LoS channel distributions. However, it is observed that the variance of the sum rate is much larger for models that experience covariate shift. Consequently, although average performance is largely maintained, the worst-case performance deteriorates when covariate shift occurs. Additionally, it is shown that a model trained on all three distributions can provide a good trade-off in terms of performance on all distributions. However, it is to be noted that the differences between the three examined channel distributions are limited. Taking this into account, more pronounced covariate shift, such as deployment

in environments with markedly different channel statistics, could have a greater negative impact on precoder reliability. Consequently, for practical deployment, it is advised to train the model on a wide range of theoretical channel models. Once the model has been deployed, real-life channel realizations can be captured. If available, these channels should be used to fine-tune the model. This is in line with our subsequent work in [82, 83] where a GNN precoder was pre-trained on simulated data, generated from channel models, and fine-tuned on measured channels. This strategy is shown to provide good generalization to real-world systems, while requiring few measured channels. This is promising given the high cost of obtaining channel measurements.

Chapter 5

Extensions to Multi-Carrier Systems

This chapter is adapted from the following paper

> T. Feys, L. Van der Perre, G. Callebaut, and F. Rottenberg, "Learning When to Learn: Distortion-Aware GNN Precoding for Multi-Carrier Large Antenna Systems with Adaptive Precoder Selection," in *ICC 2026 - IEEE International Conference on Communications*, 2026. (accepted) [29]

In previous chapters, graph neural network (GNN)-based precoding has been explored for distortion-aware precoding. This was done for the narrow-band scenario where the channel can be considered flat. In this chapter, the extension to frequency-selective channels, and multi-carrier systems is made. In a massive multiple input multiple output (MIMO) system where no nonlinear power amplifiers (PAs) are considered, extending precoders to multi-carrier systems is trivial. Due to the linearity assumption, precoding can be applied independently at each sub-carrier. In contrast, we assume a multi-carrier system with nonlinear amplifiers. In these multi-carrier systems, the non-linearity of the PA produces inter-modulation products, creating inter-carrier interference so that the precoding problem becomes coupled across subcarriers. To tackle this, we propose a GNN-based joint precoding method that optimizes precoding vectors across all subcarriers while accounting for distortion. This addresses research question Q2.3 (In a multi-carrier system, where inter-carrier interference occurs due to the nonlinear PA functions, how can NN-based distortion-aware precoding be used/adapted to allow PAs to operate close to saturation while maintaining a high data rate?). Next to this, the

consumed power of the PAs and the proposed GNN is examined to determine the overall power reduction, to answer research question Q4.1 (How much can the PA consumed power be reduced, and does this outweigh the additional inference power consumption of the required neural network (NN)?). Note that in this chapter, only a single user is considered. This is motivated by the fact that the single-user case is typically the most stringent in terms of nonlinear PA distortion, since the distortion coherently combines at the user location when few users are present [34]. The more users, with sufficiently uncorrelated channels, are present, the more the distortion is spatially spread out. Intuitively, this corresponds to summing many distortion contributions that are approximately uncorrelated across users and, hence, do not reinforce each other in any particular spatial direction. This spatial spreading of the distortion is further amplified in the multi-carrier case, when the channels across different sub-carriers have low correlation. This leads to additional, approximately uncorrelated distortion components (across sub-carriers) being summed. Consequently, when considering frequency-selective channels, the problem of PA distortion is most pronounced in the single-user case. Another stringent operating regime that should be explored in future work is the multi-user scenario where the users are sufficiently correlated in space. In this scenario, the distortion terms have a higher correlation and, as such, tend to combine constructively.

5.1 Introduction

Many works investigate how to overcome nonlinear PA distortion to improve the energy efficiency of the PAs [49, 48, 44, 55, 28]. An interesting avenue is to incorporate knowledge of the distortion into the precoder design [44, 55, 28]. This allows for the spatial suppression of the distortion in the user directions. These precoders have been designed under frequency-flat/single-carrier channel assumptions. Similarly, precoder design, assuming a theoretically ideal front-end, is done for a single flat carrier. NN-based precoders have been developed for multi-carrier systems [85]. However, spectral leakage is neglected as only linear components are considered. In [86], the performance of a massive MIMO system is evaluated in the multi-carrier case, where non-linear digital-to-analog converters are present, causing spectral leakage. However, no distortion-aware precoding techniques were proposed. In this chapter, we extend the distortion-aware precoders from Chapter 3 to the multi-carrier case by introducing a GNN which jointly computes the precoding vectors for all subcarriers while taking distortion into account. It is shown that this joint precoding delivers stable performance under varying degrees of frequency selectivity, as compared to distortion-aware precoders

developed for narrowband channels. Additionally, it is shown that the gain with respect to the classical maximum ratio transmission (MRT) precoder varies depending on the frequency-selectivity of the channel. Additionally, the consumed power of both the PAs and GNN inference is computed for a number of fixed rate constraints and varying levels of frequency-selectivity. The results highlight the ability of the GNN to enable a significant power consumption reduction, which is most prominent under flat and mildly frequency-selective channels.

5.2 Problem Formulation

5.2.1 System Model

We consider a single user downlink system with M antennas at the base station (BS), one single antenna user, and Q subcarriers. The precoded vector at subcarrier q is denoted as

$$\boldsymbol{x}_q = \boldsymbol{w}_q s_q \tag{5.1}$$

where $\boldsymbol{w}_q \in \mathbb{C}^{M\times 1}$ is the precoding vector at carrier q and $s_q \in \mathbb{C}$ is the user symbol at carrier q. Before applying the PA nonlinearity, the inverse discrete Fourier transform (IDFT) is applied to obtain the time domain signal

$$\tilde{x}_m[n] = \mathcal{F}^{-1}(x_{m,q}) = \frac{1}{\sqrt{Q}} \sum_{q=0}^{Q-1} x_{m,q} e^{\jmath 2\pi \frac{qn}{Q}} \tag{5.2}$$

where $\mathcal{F}^{-1}(\cdot)$ represents the IDFT and the time-domain signal is indicated as $\tilde{x}_m[n]$ with n being the discrete time-domain index. Afterward, the time domain signal is amplified as $\tilde{y}_m[n] = \phi_m(\tilde{x}_m[n])$, where $\phi_m(\cdot)$ denotes the nonlinear transformation caused by the PA at antenna m. The PA model used in this chapter is outlined in Section 5.2.3. The amplified signal is transformed via the discrete Fourier transform (DFT) to analyze it in the frequency domain

$$y_{m,q} = \mathcal{F}(\tilde{y}_m[n]) = \frac{1}{\sqrt{Q}} \sum_{n=0}^{Q-1} \tilde{y}_m[n] e^{-\jmath 2\pi \frac{qn}{Q}}. \tag{5.3}$$

After this, we can resume processing on a per-subcarrier basis and recover the received signal at subcarrier q as

$$r_q = \boldsymbol{h}_q^{\mathsf{T}} \boldsymbol{y}_q + \boldsymbol{v}_q \tag{5.4}$$

where $\boldsymbol{h}_q \in \mathbb{C}^{M\times 1}$ is the channel vector at subcarrier q and $\boldsymbol{v}_q$ is additive white Gaussian noise (AWGN). The full channel and precoding weights across all subcarriers are defined as $\boldsymbol{H}, \boldsymbol{W} \in \mathbb{C}^{Q\times M}$.

5.2.2 Channel Models

The wireless channel $\boldsymbol{h}_q$ is modeled by one of the following three models. Note that the channel coefficients are generated independently across antennas (independently and identically distributed (i.i.d.) per antenna–user pair), neglecting spatial correlation.

I.i.d. Rayleigh Fading per Sub-carrier

For extreme frequency-selective fading, the channel at each subcarrier q is modeled as

$$[\boldsymbol{h}_q]_m \sim \mathcal{CN}(0,1), \tag{5.5}$$

where the channel at each carrier is i.i.d.. Hence, there is no correlation in the frequency domain.

Frequency Flat Channels

As a complete opposite, frequency flat channels can be considered. These channels are modeled as

$$\boldsymbol{h}_q = \boldsymbol{h}\ \forall q \tag{5.6}$$

with $[\boldsymbol{h}]_m \sim \mathcal{CN}(0,1)$.

Rayleigh Tapped Delay Line Model

Channels with a controllable amount of correlation in the frequency domain are obtained using a Rayleigh tapped delay line (TDL) model. The time domain channel impulse response between antenna m and the user is modeled as

$$\boldsymbol{h}_m^{(t)} = \boldsymbol{C}_{\boldsymbol{h}}^{1/2}\boldsymbol{g}_m \tag{5.7}$$

$$\boldsymbol{g}_m \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_L). \tag{5.8}$$

Here $\boldsymbol{h}_m^{(t)} \in \mathbb{C}^{L\times 1}$ consists of L uncorrelated taps with an exponential decaying power delay profile (PDP), hence the covariance matrix is $\boldsymbol{C_h} = \text{diag}\left(\lambda_h^0, \cdots, \lambda_h^{L-1}\right)$ and $\lambda_h^l = e^{-\beta l}$, where β controls the amount of frequency selective fading (high β results in flatter channels, while a low β results in more frequency selectivity). The PDP of tap l is defined as $\mathbb{E}\left(\left|h_l^{(t)}\right|^2\right) = \lambda_h^l$. Additionally, to normalize the power, we set $\text{tr}\,(\boldsymbol{C_h}) = 1$. To get the channel frequency response, the DFT is applied as $\boldsymbol{h}_m = \boldsymbol{F}\boldsymbol{\Sigma}\boldsymbol{h}_m^{(t)}$, where $\boldsymbol{F} \in \mathbb{C}^{Q\times Q}$ is the DFT matrix and $\boldsymbol{\Sigma} = \left[\begin{array}{cc} \mathbf{I}_L & \mathbf{0}_{(Q-L)\times L} \end{array}\right]^{\mathsf{T}}$ is a zero padding matrix, which appends $Q-L$ zeros to $\boldsymbol{h}_m^{(t)}$.

5.2.3 Power Amplifier Model

As outlined in the system model, the amplified signal at antenna m is denoted as $\tilde{y}_m[n] = \phi_m(\tilde{x}_m[n])$, writing this in terms of amplitude modulation to amplitude modulation (AM/AM) and amplitude modulation to phase modulation (AM/PM) distortion gives

$$\tilde{y}_m[n] = \phi_{A,m}\left(\tilde{x}_m[n]\right) e^{\jmath\phi_{\varphi,m}(\tilde{x}_m[n])+\angle\tilde{x}_m[n]}. \tag{5.9}$$

In this chapter, a slightly different formulation of the modified Rapp model is considered as compared to Chapter 3. The AM/AM and AM/PM distortion for the classical modified Rapp model, as used in Chapter 3, is defined as [60]

$$\phi_A(\tilde{x}_m[n]) = \frac{|\tilde{x}_m[n]|}{\left(1+\left|\frac{\tilde{x}_m[n]}{\sqrt{p_{\text{sat}}}}\right|^{2S}\right)^{\frac{1}{2S}}} \tag{5.10}$$

$$\phi_\varphi(\tilde{x}_m[n]) = \frac{A|\tilde{x}_m[n]|^\kappa}{1+\left|\frac{\tilde{x}_m[n]}{B}\right|^\kappa}. \tag{5.11}$$

By scaling the saturation power of the PA p_{sat}, the desired average input back-off (IBO) can be set according to $\text{IBO} = p_{\text{in}}/p_{\text{sat}}$, with p_{in} being the average input power at each PA. Given that the IBO is defined as $\text{IBO} = p_{\text{in}}/p_{\text{sat}} \leq 1$, the IBO value in dB is always negative. Consequently, throughout this thesis when referring to 'larger IBO' or 'more back-off', this means a larger absolute IBO value (i.e., more negative in dB), which corresponds to operating farther from saturation. Typically, when the PA is operated far enough from saturation, no nonlinear distortion is present. However, given the classical definition of the AM/PM distortion of the modified Rapp model (5.11), the AM/PM distortion is

independent of p_{sat} and consequently of the IBO. This results in a fixed AM/PM distortion, which is also present far away from saturation. This is not consistent with the physical behavior of PAs. To correct this, a minor adjustment to the modified Rapp model is made to allow for the scaling of the AM/PM distortion with the IBO. Hence, the AM/PM distortion used in this chapter is defined as

$$\phi_{\varphi}(\tilde{x}_m[n]) = \frac{A\left|\frac{\tilde{x}_m[n]}{B}\right|^{\kappa}}{1+\left|\frac{\tilde{x}_m[n]}{B}\right|^{\kappa}}, \text{with } B = \sqrt{p_{\text{sat}}}. \tag{5.12}$$

Next to this, the modified Rapp model coefficients are set as follows[1]: $S = 2$, $\kappa = 4$, $A = -0.315$.

5.2.4 Achievable Rate

The Bussgang decomposition at the receiver is used on a per subcarrier basis to obtain the signal-to-noise-and-distortion ratio (SNDR). This is done by decomposing the received signal at subcarrier q as $r_q = B_q s_q + d_q + v_q$, where $s_q \sim \mathcal{CN}(0,1)$ is the intended user symbol at subcarrier q, d_q captures the nonlinear distortion at subcarrier q and v_q is AWGN with variance σ_v^2. The Bussgang gain at subcarrier q is $B_q = \mathbb{E}\left(r_q s_q^*\right)/\mathbb{E}\left(s_q s_q^*\right)$ [43]. Due to the Bussgang decomposition, the distortion d_q, noise v_q and intended signal s_q are uncorrelated, hence the distortion term is $\mathbb{E}\left(|d_q|^2\right) = \mathbb{E}\left(|r_q|^2\right) - |B_q|^2 p_q - \sigma_v^2$, where $p_q = \mathbb{E}\left(s_q s_q^*\right) = 1$. Consequently, the SNDR at subcarrier q is

$$\text{SNDR}_q = \frac{|B_q|^2 p_q}{\mathbb{E}\left(|d_q|^2\right) + \sigma_v^2}. \tag{5.13}$$

An achievable rate at subcarrier q, R_q, i.e., a lower bound on the capacity, is computed by considering jointly Gaussian distributed noise and distortion, which are independent from the data symbols. This can be seen as a worst case, resulting in

$$R_q = \log_2(1 + \text{SNDR}_q). \tag{5.14}$$

Additionally, the average rate across all subcarriers is defined as $R_{\text{avg}} = \frac{1}{Q}\sum_{q=0}^{Q-1} R_q$.

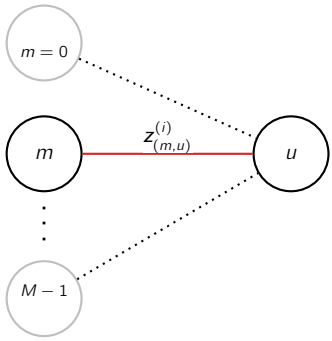


Figure 5.1: Graph representing the wireless system. Antenna nodes are on the left, the user node is on the right. Each edge contains an edge feature $z^{(i)}_{(m,u)}$.

5.3 GNN Based-precoding

As outlined in Chapter 3, GNN-based precoding is preferred over other NN architectures due to its permutation equivariance properties, which are beneficial for the precoding task. The GNN, $f(\cdot;\boldsymbol{\theta})$, can be seen as a NN that learns a mapping from the channel vector to the precoding vector over a graph, where $\boldsymbol{\theta}$ are the learnable parameters. The graph considered in this chapter has a node for each transmit antenna and a node for the user. Additionally, each transmit antenna is connected to the user node through an edge, as seen in Figure 5.1. The graph is defined as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \mathcal{V}_M \cup \{u\}$ is the set of nodes, with $\mathcal{V}_M$ being the set of antenna nodes and u being the user node, and $\mathcal{E}$ is the set of edges, which contains all edges between the antenna and user nodes i.e., $(m, u) \in \mathcal{E}\ \forall m \in \mathcal{V}_M$. The graph is undirected, meaning that $(m, u) \in \mathcal{E} \leftrightarrow (u, m) \in \mathcal{E}$, as is the physical channel between the user and antenna. A general GNN performs I message-passing iterations/layers, which update the hidden representations of the graph edges. For the considered graph, at layer i, a hidden representation is updated for each edge, denoted by $\boldsymbol{z}^{(i)}_{(m,u)} \in \mathbb{R}^{d_i}$. A single GNN layer that updates the representation of edge (m, u) is defined as

$$\boldsymbol{z}^{(i+1)}_{(m,u)} = \sigma\left(\boldsymbol{W}^{(i)}_{\text{edge}}\boldsymbol{z}^{(i)}_{(m,u)} + \boldsymbol{W}^{(i)}_{u}\boldsymbol{m}^{(i)}_{\mathcal{N}(u)}\right) \tag{5.15}$$

where $\boldsymbol{z}^{(i+1)}_{(m,u)}$ denotes the hidden representation of the edge (m, u) at layer/iteration $i + 1$ of the GNN, $\sigma(\cdot)$ is a non-linear activation function and the matrices $\boldsymbol{W}^{(i)}_{\text{edge}}, \boldsymbol{W}^{(i)}_{u} \in \mathbb{R}^{d_{i+1}\times d_i}$ are learnt using a form of stochastic gradient descent.

[1] Adapted from [58] to a PA with unit gain.

Note that the dimensions of these matrices, d_{i+1} and d_i, determine the number of features in layer $i+1$ and i respectively. The aggregation operation is defined as

$$\boldsymbol{m}_{\mathcal{N}(u)}^{(i)} = \frac{1}{M} \sum_{m' \in \mathcal{N}(u)} \boldsymbol{z}_{(m',u)}^{(i)}. \tag{5.16}$$

Depending on the setting (single-carrier/multi-carrier) the input and output edge features of the GNN differ, as outlined in the next subsections.

5.3.1 Frequency-flat Neural Network Precoder

In Chapter 3, GNN-based distortion-aware precoding was developed for the frequency-flat/single-carrier case. Even though this GNN is trained in the single-carrier case, it can still be applied in a multi-carrier system. For this, two options are available. A first option is to apply the GNN-precoder independently at each subcarrier, to obtain all precoding vectors

$$\boldsymbol{w}_q = f(\boldsymbol{h}_q; \boldsymbol{\theta}) \ \forall q. \tag{5.17}$$

Hence, the GNN is applied Q times. This ignores the inter-carrier interference due to the non-linear amplifiers. This method is referred to as GNN_1. A second option, referred to as $\text{GNN}_{q_{\text{mid}}}$, is to compute the precoding vector only for the middle carrier using the GNN as

$$\boldsymbol{w}_{q_{\text{mid}}} = f(\boldsymbol{h}_{q_{\text{mid}}}; \boldsymbol{\theta}) \tag{5.18}$$

where $q_{\text{mid}} = \lfloor \frac{Q-1}{2} \rfloor$. This precoding vector is then applied at all subcarriers. This method is preferred when the channel is fully flat across all subcarriers, as the complexity is reduced by a factor Q as compared to the previous method. For each option, the input feature for each edge is the wireless channel coefficient between the antenna and the user node at subcarrier q. The output feature for each edge is the precoding coefficient. Hence, the input and output for edge (m, u) of the network are defined as $\mathbf{z}_{(m,u)}^{(0)} = \left[\Re\{[\mathbf{h}_q]_m\}, \Im\{[\mathbf{h}_q]_m\}\right]^{\mathsf{T}} \in \mathbb{R}^{2\times 1}$, $\mathbf{z}_{(m,u)}^{(l-1)} = \left[\Re\{[\hat{\mathbf{w}}_q]_m\}, \Im\{[\hat{\mathbf{w}}_q]_m\}\right]^{\mathsf{T}} \in \mathbb{R}^{2\times 1}$. In this case, the GNN only considers a single (flat) carrier at a time, which is represented by one complex coefficient. Hence, the size of the input and output features is $d_0 = d_{l-1} = 2$, i.e., the real and imaginary parts of the channel and precoding coefficients.

5.3.2 Frequency-selective Neural Network Precoder

In the frequency-selective/multi-carrier case we consider the full channel matrix $\boldsymbol{H} \in \mathbb{C}^{Q\times M}$, which consists of Q subcarriers, each with its channel vector $\boldsymbol{h}_q \in \mathbb{C}^{M\times 1}$. The function to learn is $\boldsymbol{W} = f(\boldsymbol{H};\boldsymbol{\theta})$, where $\boldsymbol{W} \in \mathbb{C}^{Q\times M}$ is a matrix combining the precoding vectors of all subcarriers. This joint precoding over all subcarriers is denoted as GNN_Q. In this case, the input features for edge (m,u) are defined as $\boldsymbol{z}^{(0)}_{(m,u)} = [\Re\{\boldsymbol{h}_m\}, \Im\{\boldsymbol{h}_m\}]^\intercal \in \mathbb{R}^{2Q\times 1}$, where the vector $\boldsymbol{h}_m \in \mathbb{C}^{Q\times 1}$ contains the channel coefficients for all subcarriers between antenna m and the user. Similarly the output features at edge (m,u) are defined as $\boldsymbol{z}^{(l-1)}_{(m,u)} = [\Re\{\boldsymbol{w}_m\}, \Im\{\boldsymbol{w}_m\}]^\intercal \in \mathbb{R}^{2Q\times 1}$, where the vector $\boldsymbol{w}_m \in \mathbb{C}^{Q\times 1}$ contains the precoding coefficients for all subcarriers between antenna m and the user. Consequently, the size of the input and output features is $d_0 = d_{l-1} = 2Q$, i.e., the real and imaginary parts of the channel at each subcarrier. Comparing this to the case where GNN-based precoding is applied per subcarrier, the input and output feature sizes are scaled by Q.

5.3.3 Loss Functions

In the single-carrier case, the GNN is trained to maximize the expected achievable rate at one carrier as

$$\max_{\boldsymbol{\theta}} \ \mathbb{E}_{\boldsymbol{h}_q}\left(R_q(\boldsymbol{h}_q)\right), \ \text{s.t. } p_q \leq M, \tag{5.19}$$

where the expectation is taken over the channel distribution. Conversely, in the multi-carrier case, the GNN is trained to maximize the expected average achievable rate jointly across all subcarriers, subject to a power constraint. This is formulated as

$$\max_{\boldsymbol{\theta}} \ \mathbb{E}_{\boldsymbol{H}}\left(R_{\text{avg}}\left(\boldsymbol{H}\right)\right), \ \text{s.t. } p_q \leq \frac{P_{T,Q}}{Q} \tag{5.20}$$

where $p_q = \sum_{m=0}^{M-1} \mathbb{E}\left(|x_{m,q}|^2\right)$ is the power at carrier q. The total power over all carriers is defined as $P_{T,Q} = \sum_{q=0}^{Q-1} p_q = MQ$, which is obtained by setting the per-subcarrier power to $p_q = M$. To achieve this, the precoding vector at subcarrier q is normalized as

$$\boldsymbol{w}_{q,\text{norm}} = \frac{\sqrt{p_q}}{\|\boldsymbol{w}_q\|}\boldsymbol{w}_q. \tag{5.21}$$

Given that the per-carrier power is set to $p_q = M$, and that the IDFT used to get the time domain signal, spreads the power over Q time samples, the average

power at the input of each PA is $p_{in} = \frac{\sum_{m=0}^{M-1} \mathbb{E}\left(|\tilde{x}_m[n]|^2\right)}{M} = \frac{P_{T,Q}}{QM} = 1$. We note that the average IBO over all PAs is defined as $\text{IBO} = p_{in}/p_{sat}$. Given that $p_{in} = 1$ is fixed by the power constraint, we can vary the IBO by varying p_{sat}.

During training, the loss function is evaluated numerically and the above expectations are approximated by an empirical average over a mini-batch of channel realizations drawn from the training set. For each channel in the batch, $N_s = 128$ symbol vectors are independently generated and linearly precoded using the GNN output (precoding vector or matrix). Afterwards, the precoded symbols are converted to the time-domain using the IDFT, amplified, and converted back to the frequency domain using the DFT, as outlined in Section 5.2.1. Next, the rate per subcarrier is computed according to (5.14) and averaged over all carriers to produce the final loss value.

5.4 Inference Complexity

5.4.1 Classic Precoding

As a benchmark, we consider MRT precoding. This computes the complex conjugate of the channel vector for each subcarrier, which results in QM floating point operations (FLOPs). Note that the normalization of the precoding vectors is not taken into account in the complexity analysis, as this is the same for all precoder types.

5.4.2 GNN Frequency-flat Channel

We compute the complexity of the GNN that is trained on a single carrier and then is applied on each carrier independently. At inference time, the GNN is executed for each subcarrier, i.e., Q times. First, we consider a single layer of the GNN which has an input feature size of d_{in} and an output feature size of d_{out}. For this, (5.15) has to be computed M times (i.e., for each edge), while (5.16) needs to be computed only once. This leads to the following number of FLOPs

$$\mathcal{O}^{(\text{flops})}_{d_{in},d_{out}} = \underbrace{Md_{out}(4d_{in}-1)}_{(5.15)} + \underbrace{Md_{in}}_{(5.16)}. \tag{5.22}$$

When considering a GNN with I layers and a hidden feature size of d_h we have the following structure: one input layer with $d_{in} = 2$ and $d_{out} = d_h$, $I - 2$ hidden

Table 5.1: Overview of the different precoders, including feature sizes (d_0, d_{l-1}, d_h), complexity $\mathcal{O}_c$ (according to Section 5.4), required speed S_{DSP} and inference power consumption $p_{\text{cons,GNN}}$ (according to Section 5.5.2).

Precoder	Precoding Function	$d_0 = d_{l-1}$	d_h	$\mathcal{O}_c$ [MFLOPs]	S_{DSP}[GFLOPs/s]	$p_{\text{cons,GNN}}$ [W]
GNN_Q	$f(\boldsymbol{H};\boldsymbol{\theta})$	$2Q$	256	54.53	181.77	3.6
GNN_1	$f(\boldsymbol{h}_q;\boldsymbol{\theta})\ \forall q$	2	64	101.72	339.07	6.8
$\text{GNN}_{q_{mid}}$	$f(\boldsymbol{h}_{q_{mid}};\boldsymbol{\theta})$	2	64	3.18	10.60	0.2
MRT	$(\sqrt{p_q}/\|\boldsymbol{h}_q\|)\boldsymbol{h}_q^*\ \forall q$	NA	NA	10^{-3}	NA	NA

layers with $d_{in} = d_{out} = d_h$ and one output layer with $d_{in} = d_h$ and $d_{out} = 2$. When applying this GNN to all Q subcarriers, this results in a total complexity of

$$\begin{aligned}\mathcal{O}_{GNN_1}^{(flops)} &= Q\left(\mathcal{O}_{2,d_h}^{(flops)} + (I-2)\mathcal{O}_{d_h,d_h}^{(flops)} + \mathcal{O}_{d_h,2}^{(flops)}\right)\\ &= 4MQ\left(4d_h + (I-2)d_h^2\right).\end{aligned} \tag{5.23}$$

Note that the above complexity is for the GNN trained for the single-carrier case which is then applied independently at each subcarrier of the multi-carrier system, this is denoted as GNN_1. Alternatively, we defined $\text{GNN}_{q_{mid}}$ which computes the precoder at the middle carrier and applies it at each subcarrier of the multi-carrier system. Hence, the complexity of $\text{GNN}_{q_{mid}}$ is Q times lower as compared to (5.23).

5.4.3 GNN Frequency-selective Channel

In the multi-carrier case we can reuse the computational complexity of one layer of the GNN as computed in (5.22). The GNN for joint precoding has the following structure: one input layer with $d_{in} = 2Q$ and $d_{out} = d_h$, $I-2$ hidden layers with $d_{in} = d_{out} = d_h$ and one output layer with $d_{in} = d_h$ and $d_{out} = 2Q$. Given that the GNN in the multi-carrier case jointly computes the precoders across all subcarriers the GNN only needs to be executed one time (as compared to Q times for the single-carrier case). This results in the following computational complexity

$$\begin{aligned}\mathcal{O}_{GNN_Q}^{(flops)} &= \mathcal{O}_{2Q,d_h}^{(flops)} + (I-2)\mathcal{O}_{d_h,d_h}^{(flops)} + \mathcal{O}_{d_h,2Q}^{(flops)}\\ &= 4M\left(4Qd_h + (I-2)d_h^2\right).\end{aligned} \tag{5.24}$$

The total complexity for each precoder, considering the parameters used in the simulations in Section 5.6 are given in Table 5.1.

5.5 Power Consumption

5.5.1 PA Consumption

For a class B amplifier, the following model of the energy efficiency can be used [73]. The efficiency of PA m is defined as

$$\eta_m = \eta_{\max}\sqrt{\frac{p_m}{p_{\text{sat}}}}. \tag{5.25}$$

Here p_{sat} is the saturation power of the PA, while $\eta_{\max}$ is the maximum achievable efficiency, for a class B amplifier $\eta_{\max} = 0.785$ [73]. Next to this, p_m is defined as the power at the output of amplifier m

$$p_m = \mathbb{E}\left(|\tilde{y}_m[n]|^2\right) = \mathbb{E}\left(|\phi(\tilde{x}_m[n])|^2\right). \tag{5.26}$$

The total consumed PA power is then given by

$$p_{\text{cons,PA}} = \sum_{m=0}^{M-1}\frac{p_m}{\eta_m} = \frac{\sqrt{p_{\text{sat}}}}{\eta_{\max}}\sum_{m=0}^{M-1}\sqrt{\mathbb{E}\left(|\tilde{y}_m[n]|^2\right)}. \tag{5.27}$$

From this expression, it can be seen that the consumed power is not proportional to the average input back-off but rather to the sum of the square root of the output power of the PAs. Note that $\mathbb{E}\left(|\tilde{x}_m[n]|^2\right)$ and $\mathbb{E}\left(|\tilde{y}_m[n]|^2\right)$ represent dimensionless digital powers. These can be translated to an actual power value by scaling it with a physical reference power (in watt) as $p_{\text{ref},m}\mathbb{E}\left(|\tilde{x}_m[n]|^2\right)$. Note that in this thesis $p_{\text{ref},m} = 1$ W is assumed. This is inline with sub-6 GHz BSs such as the one in [74], where the power per antenna ranges from 1.25 to 3.125 watts, depending on the exact frequency band.

To clarify some of the PA power consumption results in Section 5.6 we will outline a relation between the PA input power allocation and the PA consumed power. Let the PA input power at antenna m be defined as $p_m^{\text{in}} = \mathbb{E}\left(|\tilde{x}_m[n]|^2\right)$. In vector form, this is denoted as $\boldsymbol{p}^{\text{in}}$. The total input power to all PAs is then $\sum_{m=0}^{M-1} p_m^{\text{in}} = M$. The PA output power at antenna m can then be written as a mapping from the PA input power as

$$p_m = \mathbb{E}\left(|\phi(\tilde{x}_m[n])|^2\right) = g(p_m^{\text{in}}) \tag{5.28}$$

where $g(\cdot)$ is determined by $\phi_A(\cdot)$ and the statistics of $\tilde{x}_m[n]$. We can rewrite the class B efficiency model as

$$p_{\text{cons,PA}} = \frac{\sqrt{p_{\text{sat}}}}{\eta_{\max}}\sum_{m=0}^{M-1}\sqrt{p_m} = C\sum_{m=0}^{M-1}\psi(p_m^{\text{in}}) \tag{5.29}$$

where $\psi(a) := \sqrt{g(a)}$ and $C = \frac{\sqrt{p_{\text{sat}}}}{\eta_{\max}}$. We propose the following proposition.

Proposition 1. *If $\psi(\cdot)$ is concave on $\mathbb{R}_+$, then for any input power allocation $\boldsymbol{p}^{\text{in}}$ with the same total power $\sum_m p_m^{\text{in}} = M$,*

$$p_{\text{cons,PA}} = C \sum_{m=0}^{M-1} \psi\left(p_m^{\text{in}}\right) \leq CM\psi\left(1\right), \tag{5.30}$$

with equality iff $p_m^{\text{in}} = 1 \ \forall m$.

Proof. Given the concavity of $\psi(\cdot)$ and Jensen's inequality we have

$$\frac{1}{M}\sum_m \psi\left(p_m^{\text{in}}\right) \leq \psi\left(\frac{1}{M}\sum_m p_m^{\text{in}}\right) = \psi(1). \tag{5.31}$$

By multiplying both sides by CM we prove Proposition 1. □

Thus, for a fixed total input power, any non-uniform (more spread) input power allocation achieves a smaller total consumed PA power, given that $\psi(\cdot)$ is concave. Note that the analytical proof of the concavity of $\psi(\cdot)$ is non-trivial due to the expectation operator in $g(\cdot)$. Thus, the concavity was verified numerically by verifying that $\psi''(a) \leq 0$ for the Rapp PA model and input statistics used in this chapter. Intuitively, the results of Proposition 1 can be explained as follows: The more non-uniform the power allocation, the more certain antennas are saturated. These saturated antennas have a higher efficiency, while they also consume more of the power budget as compared to the less efficient non-saturated antennas. This results in an overall reduction in the consumption of the PAs. In Section 5.6, as a measure of non-uniformity or spread, the Herfindahl–Hirschman index (HHI) (which is directly linked to the squared L2-norm) is used, which is defined as

$$\text{HHI} = \sum_{m=0}^{M-1}\left(\frac{p_m^{\text{in}}}{\sum_{m'} p_{m'}^{\text{in}}}\right)^2 = \sum_{m=0}^{M-1}\left(\frac{p_m^{\text{in}}}{M}\right)^2 = \frac{1}{M^2}\|\boldsymbol{p}^{\text{in}}\|_2^2. \tag{5.32}$$

This metric always falls in the range $1/M \leq \text{HHI} \leq 1$, where it is equal to $1/M$ if the power is allocated uniformly and equal to 1 if all power is allocated to a single antenna. Additionally, up to a constant scaling, the HHI corresponds to the second-order spatial moment of the power allocation across antennas. Given Proposition 1, the more non-uniform the power allocation is, i.e., the higher the HHI-score, the lower the PA consumed power will be.

5.5.2 GNN Inference Consumption

In this section, the consumed power of the hardware accelerator needed to run the GNN is quantified. Given the computational complexity of the proposed precoders in Table 5.1, the worst case complexity is that of GNN_1 with $\mathcal{O}_c = 101.72$ MFLOPs. To establish the desired speed (FLOPs/s) of the hardware accelerator, the channel coherence time is computed to determine the time available to compute the precoding coefficients. The coherence time is inversely proportional to the maximum Doppler frequency $f_m = \frac{v}{c} f_c$ and is often defined as $T_c = \frac{1}{2f_m}$ [19]. Assuming a carrier frequency of $f_c = 5$ GHz and a receiver velocity of $v = 10$ m/s, this gives a coherence time of 3 ms. As a design choice, it is assumed that 10% of the coherence time can be used to compute the precoding coefficients. For instance, in [75, 76], a latency margin of 153 µs is available for precoding within a 0.5 ms radio slot (supporting a minimum coherence time of 430 µs), equating to 35% of the coherence time that is available for precoding. Given our larger coherence time of 3 ms, a conservative 10% is adopted. This establishes a required hardware accelerator speed of $S_{\text{DSP}} = 339.07$ giga floating point operations per second (GFLOPs/s). Based on [69] the state-of-the-art Manticore accelerator [77] is selected. The full Manticore design consists out of 4096 cores, however, in [77] a prototype with 24 cores was developed and measured. This prototype is able to achieve 50 GFLOPs/s with a power consumption of 1 W. Given the parallel architecture of the GNN it can be deployed across multiple cores. Hence, in order to reach the desired speed of 339.07 GFLOPs/s, 163 cores can be utilized rather than 24. This achieves the desired speed and precision as it enables double precision[2] at an efficiency of $\eta_{\text{DSP}} = 50$ GFLOPs/s/W. Based on this efficiency, the inference power consumption of the GNN is computed according to

$$p_{\text{cons,GNN}} = \frac{S_{\text{DSP}}}{\eta_{\text{DSP}}}. \tag{5.33}$$

The inference power consumption for the different GNN-based precoders is illustrated in the final column of Table 5.1. Note that the required inference speed and consumption of MRT is assumed to be zero, as it is negligible compared to that of the GNN precoders.

[2]Note that the GNN requires only single precision (FP32), while the accelerator can deliver double precision (FP64).

5.6 Simulation-based Performance Assessment

For the simulations, $M = 32$ transmit antennas and $Q = 32$ subcarriers are considered. The linear PA gain is set to one. The dataset for the GNN-based precoders consists of a training set of 500.000 channels, a validation set of 1000 channels and a test set of 10.000 channels. For GNN_Q the channels are generated according to the TDL channel model with a β value that is sampled from 50 possible values logarithmically spaced between 10^{-2} and 10^2. The models are trained for 30 epochs, using the Adam optimizer with a learning rate of 0.5×10^{-3} and batch size of 64. Training is done at $\text{SNR}_q = P_{T,Q}/Q\sigma_v^2 = 20$ dB. GNN_1 and $\text{GNN}_{q_{mid}}$ consist of $I = 8$ layers with a hidden feature size of $d_h = 64$. GNN_Q consists of $I = 8$ layers with a hidden feature size of $d_h = 256$. For both models, a leaky ReLu activation is used. Additionally, the output of the model is normalized according to (5.21) in order to respect the power constraint.

5.6.1 Results Fixed IBO

In this subsection, the different precoders are analyzed at a fixed IBO = -3 dB. This operates the PAs significantly closer to their saturation point as compared to traditional operating points that target linearity. To evaluate the performance of the different precoders, different channel conditions are considered. First, two extreme cases are considered in Figure 5.2, where fully frequency flat and fully frequency uncorrelated (i.i.d.) channels are considered for a fixed IBO$= -3$ dB. When channels are considered that are i.i.d. across sub-carriers, MRT has the best performance as compared to the GNN-based precoders. When the channels are i.i.d. across sub-carriers, the distortion is more uniformly distributed in space, due to the addition of many uncorrelated components (subcarriers), hence the distortion does not receive a considerable beamforming gain. This is a similar phenomenon to the case where many users are present [34]. In this case, the non-linear distortion is more spatially spread out due to the addition of many uncorrelated components (users). When frequency flat channels are considered, the distortion is beamformed into the user direction, leading to a lower performance of MRT. Under these channel conditions the GNN-based precoders can achieve a considerable increase in rate of 5.3 bits/ channel realization at an $\text{SNR}_q = 30$ dB. This rate increase is achieved by beamforming the distortion in non-user directions [28]. Notably, we see that under flat channel conditions, all three GNN schemes achieve approximately the same performance.

Second, the TDL channel model is considered, where the amount of frequency selectivity is tuned by changing β. We note that for this channel model, by

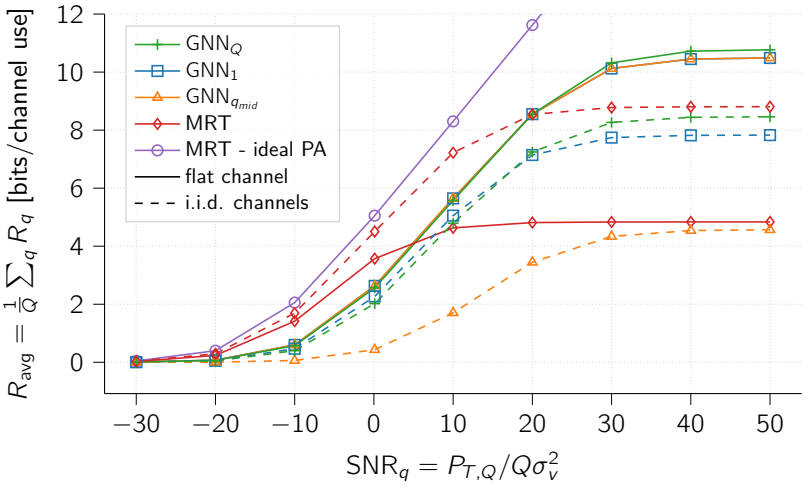


Figure 5.2: Rate vs SNR_q averaged over the channels of the test set. For $M = 32, Q = 32$ and IBO $= -3$ dB and $\text{SNR}_q = 20$ dB during training, on strongly frequency selective/i.i.d. (dashed) and frequency flat channels (full).

considering a limited number of taps $L = 10$ (with $L \ll Q$), the sub-carriers are never fully uncorrelated. In Figure 5.3, the rate is illustrated for varying degrees of frequency selectivity (β) at a fixed IBO = -3 dB. When β is large, the channel can be considered flat, resulting in the GNN-based precoders outperforming MRT. As β decreases, the channel becomes more frequency selective, leading to a decreased performance of GNN_1 and $\text{GNN}_{q_{mid}}$ as these NNs are trained for the single-carrier/flat case. Additionally, as β decreases, the performance of MRT improves, due to the non-linear distortion becoming more spatially spread out. By jointly precoding all subcarriers, as done by GNN_Q, we observe that the performance is maintained, even for small values of β. Notably, even for very low β values GNN_Q outperforms MRT. This is because the sub-carriers are never fully uncorrelated, even for $\beta = 0$, due to the limited number of considered taps.

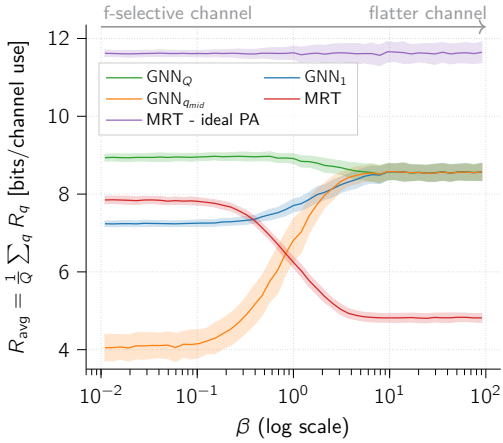


Figure 5.3: Rate vs β with ± 1 Standard Deviation averaged over the channels of the test set. For $M = 32$, $Q = 32$, $\text{SNR}_q = 20$ dB and IBO $= -3$ dB, on a channel with varying degrees of frequency selectivity.

5.6.2 Results Varying IBO

In this subsection, the different precoders are assessed for different IBO settings, namely IBO $\in \{0, -3, -6, -9, -12, -15, -18\}$ dB. Note that at each IBO value the GNN precoders are retrained. In Figure 5.4, the rate as well as the consumed PA power (according to (5.27)) are depicted as a function of the frequency-selectivity of the channel β. To avoid overloading the figure, only a subset of IBO values is shown. To interpret the results of Figure 5.4, the following paragraphs each discuss a different IBO regime (this can either be a singular IBO value or a group of IBO values if the behavior is unchanged). Within each paragraph (IBO regime), the average rate (R_{avg}) and the consumed PA power ($p_{\text{cons,PA}}$) are discussed for low and high β values.

Low IBO: -18 dB At a low β value, i.e., a frequency-selective channel, we see that MRT achieves the highest rate, while GNN$_1$ can match this rate. Conversely, GNN$_{q_{mid}}$ achieves a very poor rate as compared to MRT. This is due to the strong frequency selectivity of the channel. GNN$_{q_{mid}}$ only considers the middle subcarrier for precoding and applies the same precoder at all carriers; consequently, for strongly frequency-selective channels (low β values) it performs poorly. Looking at GNN$_Q$, we see a slight performance drop with respect to MRT. Given that at -18 dB IBO there is no significant distortion, the precoding problem becomes

close to decoupled across the subcarriers. Consequently, precoding on a per-sub-carrier basis is close to optimal in this case (as is done by MRT and GNN_1). However, GNN_Q is still trained in a manner where coupling across the sub-carriers is assumed. As such, the loss function of GNN_Q becomes overly complex, leading to a slight performance drop as compared to GNN_1. Next, we examine the PA consumed power at low β. We see that the consumed power of MRT, GNN_1 and GNN_Q is approximately the same. This is confirmed by their similar HHI-scores in Figure 5.5. We see that the consumed power of $\text{GNN}_{q_{mid}}$ is significantly lower. This is explained by the fact that $\text{GNN}_{q_{mid}}$ applies the same precoder, based on the middle carrier, at each sub-carrier. Consequently, the antennas that have a slightly higher channel gain at the middle carrier will receive more power, as by precoding based on the middle carrier only, the effect of this slightly higher channel gain is compounded. This leads to some antennas receiving more power than others, which is confirmed by the higher HHI scores in Figure 5.5. Consequently, a higher HHI-score corresponds to a more non-uniform power allocation, which, according to Proposition 1 results in a lower power consumption.

Considering high β values, i.e., channels that experience low or no frequency selectivity. The performance in terms of the rate of all four considered precoders converges to the same value. Looking at PA consumed power, we see that the consumed power of all GNNs converges to the same value, while MRT has a slightly lower consumption, which is explained by its higher HHI-value in Figure 5.5.

Low IBO: -15 dB At this IBO again we start with discussing the rate and PA consumed power for low β/frequency-selective channels. In this regime, we see that MRT achieves the highest rate, followed by GNN_1 and GNN_Q. Again, we see that the rate of $\text{GNN}_{q_{mid}}$ is deteriorated due to its inability to handle frequency selectivity. With respect to the PA consumed power, we see a similar trend as before, where the consumed power of MRT, GNN_1, and GNN_Q is approximately the same, while the consumed power of $\text{GNN}_{q_{mid}}$ is slightly lower due to its higher HHI-score that stems from the compounded power across sub-carriers at antennas that experience a slightly higher channel gain.

For high β values, we see that the GNN precoders converge to the same rate, which is higher than the rate of MRT. From Chapter 3, for the frequency-flat/single-carrier case, it is known that when a single (flat) carrier is considered and the PAs are deeply saturated (i.e., a small IBO), the GNN-based distortion-aware precoder cancels the distortion in the user direction by saturating a subset of the antennas. This behavior is similar to the closed-form Z3RO precoder as studied in Chapter 2. However, here we see a different phenomenon as the rate is improved, but no antennas are saturated as can be seen from the low HHI-score

in Figure 5.5. This is explained by the fact that the operating point of the PAs at IBO $= -15$ dB is still far from saturation. Consequently, during training the GNNs learn a different policy under these low-distortion conditions. Rather than saturating some antennas with an opposite phase shift to cancel the distortion, an opposite phenomenon is observed. The power of antennas with high channel gain is allocated to antennas with a lower channel gain, such that saturation is avoided in the high channel gain antennas. This can be seen in Figure 5.6 where antennas with a lower channel gain receive more power as compared to MRT and antennas with a higher channel gain receive less power as compared to MRT[3]. Given that at IBO $= -15$ dB the average operating point is far from saturation, this power allocation effectively avoids entering the saturation regime. As such, a reduction in array gain is effectively traded for less distortion, which results in a higher rate as compared to MRT. This is confirmed by the lower HHI-score, which suggests a more uniform power allocation. Given this lower HHI-value for the GNNs, due to Proposition 1, the consumed PA power of the GNNs should be higher than that of MRT. This is confirmed by the PA consumption in Figure 5.4.

Moderate IBO: -9 dB For an IBO of -9 dB at low β, we see that GNN_Q and MRT achieve a similar rate, which is significantly higher than that of GNN_1. Besides this, again we see the poor performance of $\text{GNN}_{q_{mid}}$ under frequency-selective channels. In terms of PA consumed power MRT, GNN_1 and GNN_Q consume approximately the same amount of power, while again the consumption of $\text{GNN}_{q_{mid}}$ is slightly lower due to its higher HHI-score.

At high β, we see the same effect as at an IBO of -15 dB, however, this time the effects are more pronounced. The rate of the GNN precoders converges to the same value, which is significantly higher than the rate of MRT. The PA power consumption of all precoders is very similar at high β, with MRT having a slightly lower power consumption due to its higher HHI-score.

High IBO: -3-0 dB In this regime, the PAs are operated very close to their saturation point. At low β GNN_Q achieves the highest rate, which is approximately 1 bit/channel use higher than the rate of MRT. In the frequency-flat case, similar gains are obtained. These gains are achieved by saturating antennas with opposite phase shifts to cancel the distortion as was illustrated in the single-carrier case in Chapter 3. When simply applying these distortion-aware precoders on a per-sub-carrier basis, as done by GNN_1, we see reduced performance in terms of rate.

[3]Note that this plot is only for GNN_Q but the same behavior is observed for GNN_1 and $\text{GNN}_{q_{mid}}$.

Given that the channel varies across carriers, GNN_1 tries to saturate different antennas at each carrier, which cancels out when combining all carriers through the DFT to generate the time domain signal. This can be seen by the low HHI-score of GNN_1 in Figure 5.5. Conversely, $\text{GNN}_{q_{mid}}$ achieves a high HHI-score, indicating saturated antennas. However, it only considers a single carrier; consequently, this results in a poor rate. Notably, we see that GNN_Q achieves a similar HHI-score to $\text{GNN}_{q_{mid}}$. This is because GNN_Q manages to saturate certain antennas (with opposite phase), across carriers, to cancel distortion, which increases its rate. With respect to PA consumed power, this saturation effect is reflected as GNN_Q and $\text{GNN}_{q_{mid}}$ achieve the lowest consumption, due to their high HHI-scores, while MRT and GNN_1 produce a higher consumption.

At high β, all GNN precoders achieve the same rate, which is significantly higher than that of MRT. As said before, this gain is achieved by saturating certain antennas. This results in a higher HHI-score, which leads to a lower PA power consumption of the GNN precoders as compared to MRT. In sum, optimizing the GNN precoders to maximize the sum rate under severe nonlinear distortion results in the saturation of certain antennas. This produces a higher HHI-score, which directly results in a reduced PA power consumption.

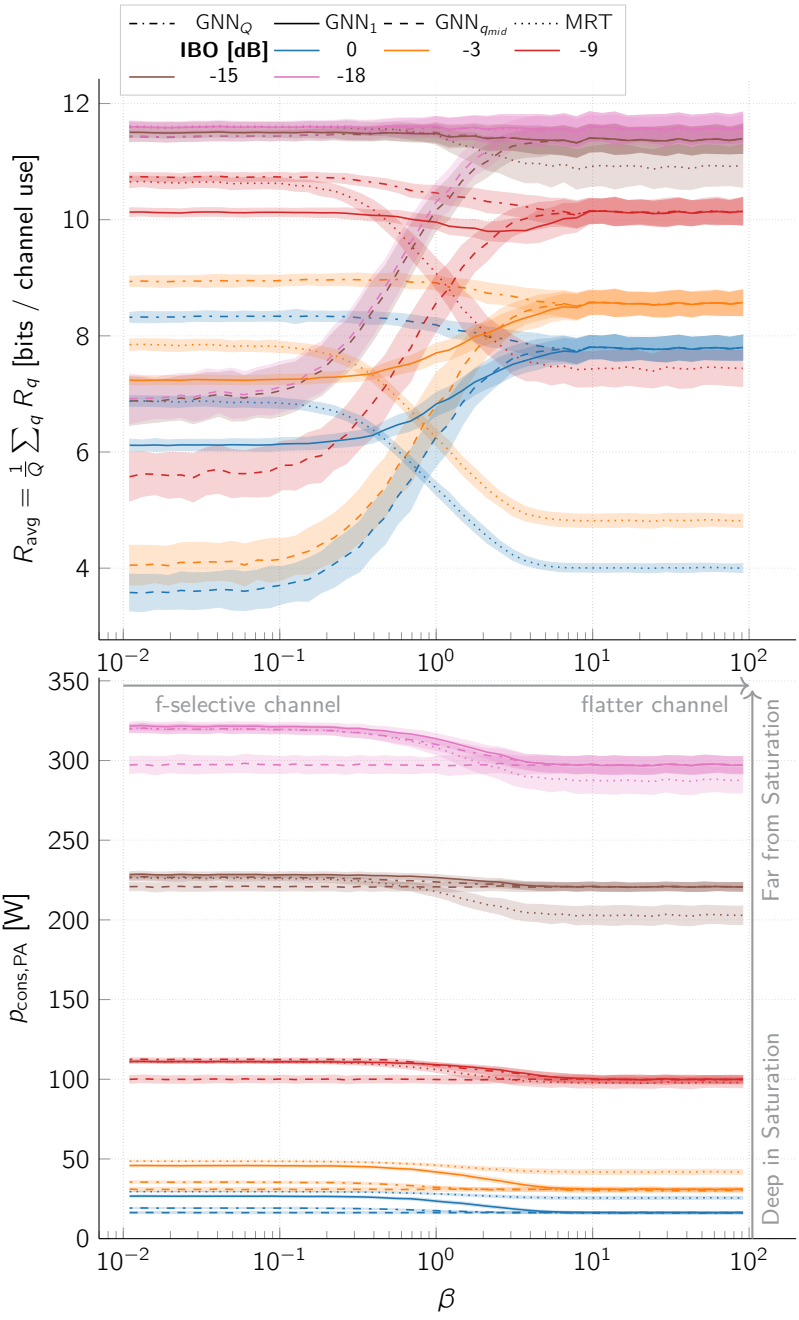


Figure 5.4: Comparing the average rate and PA consumed power of the different precoders in function of β for a number of IBO values. For $M = 32$, $Q = 32$ and $\text{SNR}_q = 20$ dB.

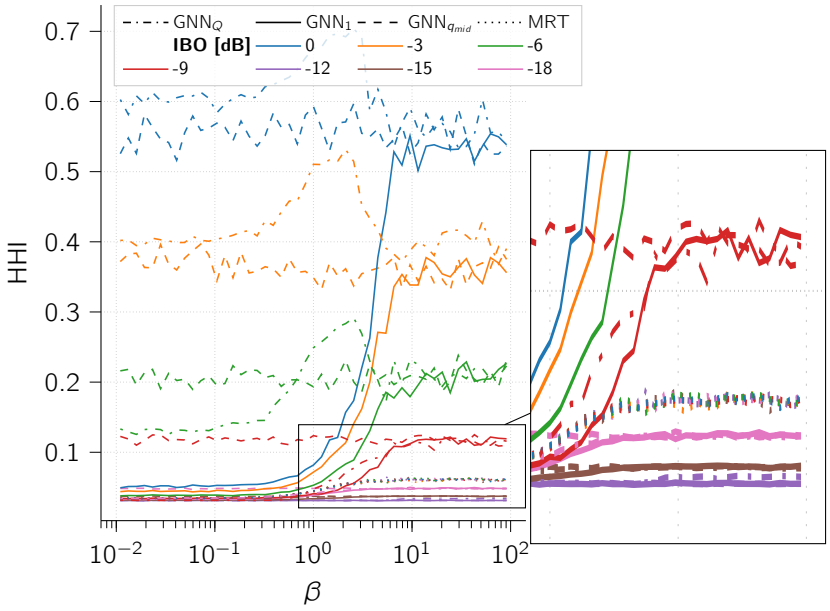


Figure 5.5: HHI-score of the different precoders at different IBOs as a function of the channel frequency-selectivity (β). When HHI $= 1/M$ the power is allocated uniformly across the antennas (implying a higher power consumption), when HHI $=$ 1 all power is allocated to a single antenna (implying a lower power consumption).

5.6.3 Power Consumption Comparison

In the previous section, the precoders were evaluated in terms of rate and PA consumed power at fixed IBO values. At these IBO points, the different precoders achieve a different rate, due to their (in)ability to handle nonlinear distortion. However, to fairly compare the PA power consumption of the different precoders, the consumption for a fixed quality of service (QoS) constraint should be considered. The methodology to make this comparison is explained in the next paragraph.

Methodology To produce a fair comparison, the aim is to compare the consumed PA powers of the different precoders, for all β values, such that a rate constraint $R \geq R_{\text{min}}$ is met. The required PA consumed power to meet this constraint is obtained as follows. First, the rate constraint value R_{min} is set. Second, for each β value and for each precoder, the lowest possible IBO value (closest to saturation) is selected such that the rate constraint is still met. This is done

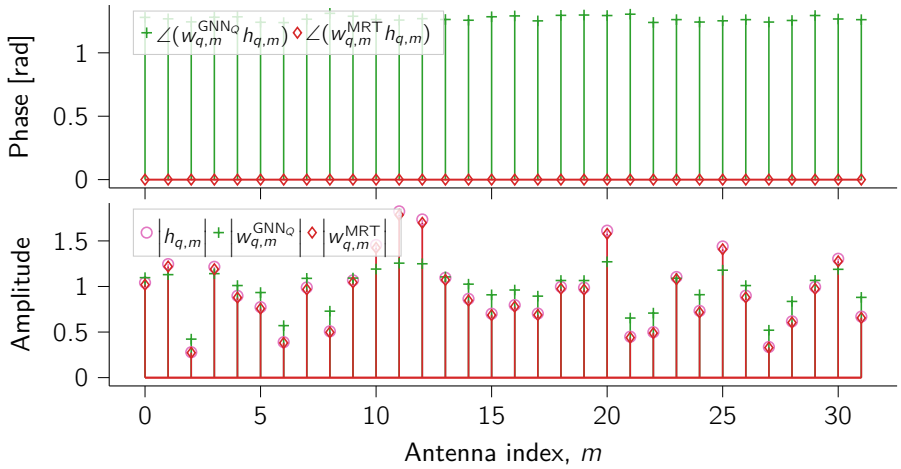


Figure 5.6: Amplitude and phase plot per antenna for a single subcarrier at IBO $= -15$ dB, $\beta = 11.5$. Comparing GNN_Q with MRT.

based on the top plot in Figure 5.4, this IBO value is denoted as $\mathrm{IBO}_{\mathrm{min}}$. Next, for each β and precoder, the PA consumed power associated to this $\mathrm{IBO}_{\mathrm{min}}$ is selected using the bottom plot of Figure 5.4. This gives the consumed PA power for each precoder, at all β values, such that they all meet the same QoS, i.e., they all meet the rate constraint. The rate constraints that are examined are $R_{\mathrm{min}} \in \{4, 8, 11\}$ bit/channel use. We note that a high rate constraint such as 11 bit /channel use exceeds what is typically expected for a single-antenna user under practical signal-to-noise ratio (SNR) conditions. However, it is included in this study as it forces operation in a regime where nonlinear PA effects are dominant.

The achievable rate constraints in bit/channel use can be translated to a spectral efficiency by accounting for excess bandwidth due to guard bands. Specifically, we assume that a fraction $\alpha_{\mathrm{rol}} = 0.1$ of the occupied bandwidth is reserved as guard-band overhead, e.g., due to pulse-shaping and spectral containment requirements. The resulting spectral efficiency is computed as $\eta_{\mathrm{SE}} = \frac{R_{\mathrm{min}}}{1+\alpha_{\mathrm{rol}}}$. This gives the following spectral efficiency constraints $\eta_{\mathrm{SE}} \in \{3.6, 7.3, 10\}$ bit/s/Hz. This can be further concretized to a bit rate constraint R_{b} by assuming a specific bandwidth. For instance a bandwidth of $B = 100$ MHz results in the following throughput constraints $R_{\mathrm{B}} \in \{363.6, 727.3, 1000\}$ Mbit/s. Note that these achievable rate, spectral efficiency, and throughput values assume a normalized large-scale fading gain of one and do not account for path loss or shadowing effects.

Results The PA power consumption comparison for different R_{min} values is illustrated in Figure 5.7, 5.8 and 5.9. For a low rate constraint, i.e., $R_{\text{min}} = 4$ bits/channel use, all precoders can easily meet the constraint. This can be seen from the IBO_{min} values in Figure 5.7b, which suggests that PAs can be operated deep in saturation ($\text{IBO}_{\text{min}} = 0$ dB for all precoders except for $\text{GNN}_{q_{mid}}$ at low β which requires $\text{IBO}_{\text{min}} = -3$ dB) while the rate constraint is still maintained. In terms of PA consumed power Figure 5.7a highlights that even under this lenient rate constraint GNN_Q has a significant reduction in consumed power of 9.5-10.6 watts with respect to MRT, depending on the channel conditions (β). This represents a $34.9 - 37.2$ % reduction in consumed PA power. Under frequency-flat conditions, GNN_1 and $\text{GNN}_{q_{mid}}$ achieve similar reductions in consumed PA power. However, as the channel becomes more frequency selective, this reduced PA consumption diminishes. For $\text{GNN}_{q_{mid}}$, this is due to the need for a higher IBO value to meet the rate constraint. While for GNN_1 the frequency selectivity of the channel results in the power allocation over the antennas becoming more uniform, which according to Proposition 1, results in a higher PA power consumption.

In Figure 5.8, a rate constraint of $R_{\text{min}} = 8$ bits/channel use is considered. From Figure 5.8b it can be seen that, depending on the precoder type and channel conditions, the operating point of the PAs can either be closer or farther from saturation in order to meet this rate constraint. It can be seen that GNN_Q operates closest to saturation across all channel conditions compared to the other precoders. GNN_1 and $\text{GNN}_{q_{mid}}$ can operate under similar levels of distortion when the channel is sufficiently flat ($\beta \geq 10$), while for frequency selective channels their PAs need to be backed off more. Across all channel conditions, MRT needs to operate the PAs the furthest from saturation to meet the rate constraint. This results in a higher PA power consumption, as can be seen in Figure 5.8a. GNN_Q requires 56-112 watts less to meet the same rate constraint as MRT. This corresponds to a reduction of $74.5 - 83.5$ % in consumed PA power. Again, we see that the GNN precoders designed for the single-carrier case (GNN_1 and $\text{GNN}_{q_{mid}}$) achieve similar power reductions for sufficiently flat channels, while this diminishes as the channel becomes more frequency selective.

Finally, in Figure 5.9 the highest rate constraint is considered at $R_{\text{min}} = 11$ bits/channel use. From Figure 5.9b it is clear that under this stringent rate constraint, even the GNN precoders require that their PAs are sufficiently backed off. For frequency-selective channel conditions, all precoders require the same IBO of $\text{IBO}_{\text{min}} = -12$ dB. When the channel is more frequency-flat, the GNN precoders require $\text{IBO}_{\text{min}} = -15$ dB while MRT requires $\text{IBO}_{\text{min}} = -18$ dB. Additionally, $\text{GNN}_{q_{mid}}$ can only meet the rate constraint for flat channels (high β). Figure 5.9a illustrates the impact on the PA consumed power. GNN_Q and GNN_1 achieve the same reduction of PA consumed power of 56-66 watts for $\beta \geq 1$, this corresponds

to a $56.5 - 66.7$ % reduction in PA consumed power. At lower β values, this power reduction disappears and PA consumed power of all precoders becomes equal.

Total Power Consumption In the previous paragraph, the gains in terms of reduction of PA consumed power were outlined. However, the additional power consumption due to the inference of the GNN precoders has not yet been considered. In Section 5.5.2, the inference complexity of the GNN precoders and their associated power consumption were analyzed, and the resulting inference power consumption is summarized in Table 5.1. Overall, the GNN inference power consumption is low ranging from 0.2 to 6.8 watt depending on the type of GNN used. Comparing this to the gains in terms of PA power reduction illustrated in the previous paragraph, which range from 9.5 to 112 W, the GNN inference power consumption is neglegible under most conditions. This is especially the case considering that the proposed joint precoding approach only consumes 3.6 watt. Consequently, the GNN inference consumption is negligible with respect to the PA power reduction under most operating conditions.

However, in certain operating regimes, for instance when considernig a very high rate constraint coupled with highly frequency-selective channels or when a very low rate constraint is considered, there are no (or very limited) gains in terms of PA power consumption. In these cases, the use of GNN-based beamformers is not advised, as it needlessly increases the digital signal processing (DSP) complexity without providing corresponding power savings.

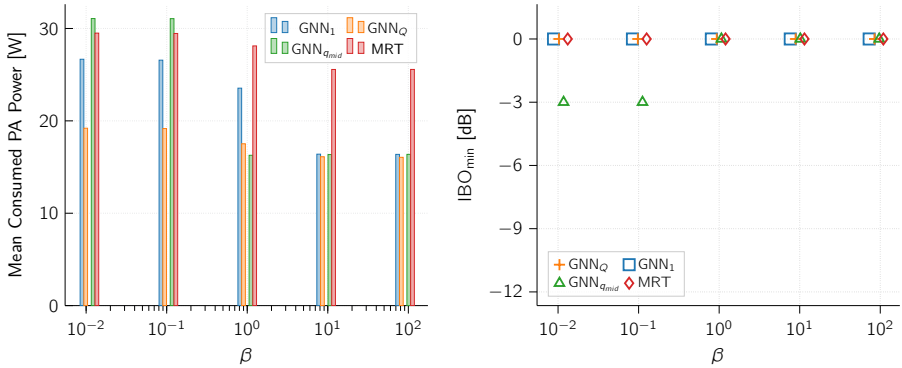


Figure 5.7: Required $p_{\text{cons,PA}}$ (a) and IBO_{max} (b) to meet $R \geq 4$ in function of β.

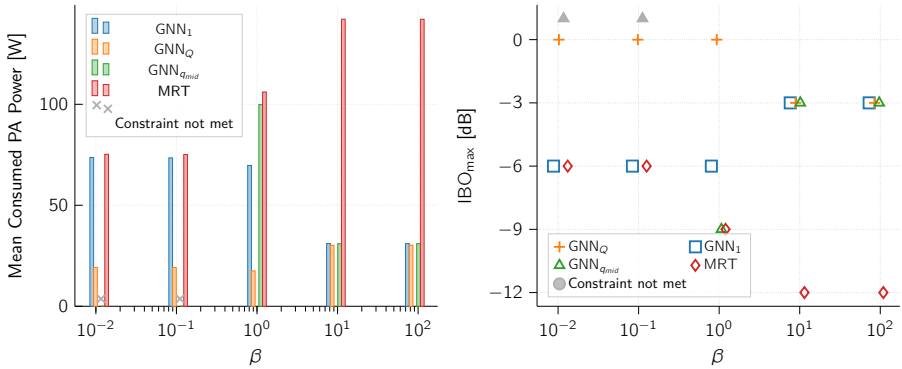


Figure 5.8: Required $p_{\text{cons,PA}}$ (a) and IBO_{max} (b) to meet $R \geq 8$ in function of β.

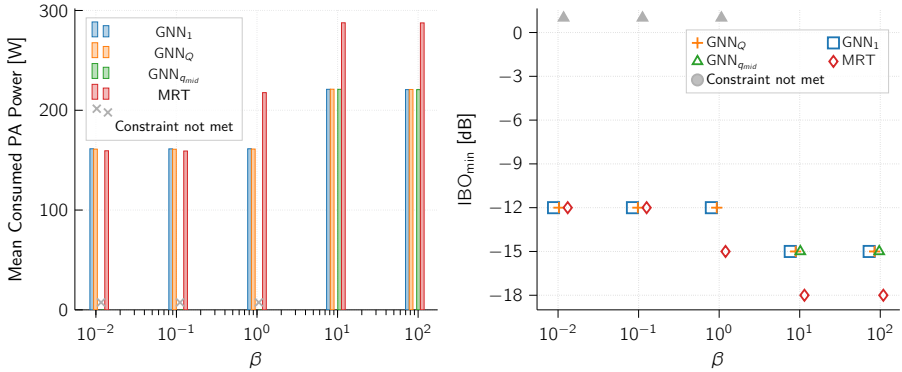


Figure 5.9: Required $p_{\text{cons,PA}}$ (a) and IBO_{max} (b) to meet $R \geq 11$ in function of β.

5.7 Out-of-band Emissions and Oversampling

In the previous and the current chapter, the main focus has been on in-band distortion/performance. Additionally, non-linear PAs create spectral regrowth outside the allocated frequency band, i.e., out-of-band emissions. These out-of-band emissions are subject to stringent regulations defined by third generation

partnership project (3GPP) and European telecommunications standards institute (ETSI), as further discussed in Chapter 7.

Previously, the nonlinearity was applied at the symbol rate; i.e., no oversampling was considered. Although oversampling is typically required in hardware implementations to characterize out-of-band spectral regrowth, it was previously not considered for simplicity and computational efficiency. As such, as a starting point, symbol rate modeling was considered. In particular, symbol-rate modeling substantially reduces simulation times, which is especially critical when training neural-network-based precoders, where each training run involves a large number of forward and backward passes. Prioritizing in-band distortion is also consistent with the main objective of this work and with the broader distortion-aware precoding literature.

Despite this simplification, the results obtained under symbol-rate modeling remain meaningful. First, neglecting oversampling can be seen as a worst-case in-band scenario, in which all spectral regrowth is aliased back into the useful signal band. The in-band distortion levels reported in previous sections should therefore be regarded as an upper bound. Second, because both the proposed GNN precoder and the baseline zero forcing (ZF)/MRT schemes are subjected to the same aliasing conditions, the comparative analysis of their robustness to nonlinear distortion remains fair.

For completeness, in this section, we evaluate the in-band and out-of-band (OOB) emissions of the GNN-based precoder when considering oversampling. Note that for this analysis, the GNN-based precoder is not retrained; i.e., it is trained without considering oversampling (unless specified otherwise) and evaluated with oversampling.

5.7.1 Oversampling

The system model presented in Section 5.2.1 operates at the symbol rate f_{sym}, i.e., the time-domain signal $\tilde{x}_m[n]$ contains exactly Q samples per orthogonal frequency division multiplexing (OFDM) symbol, matching the number of subcarriers. This critically-sampled representation is insufficient to faithfully capture the output of a nonlinear PA. The memoryless nonlinearity $\phi_m(\cdot)$ generates spectral regrowth, i.e., intermodulation products that spread beyond the original signal bandwidth $[-f_{\text{sym}}/2, f_{\text{sym}}/2]$. When sampling at the symbol rate, these out-of-band components alias back into the in-band region, which generally leads to an overestimation of the in-band distortion and, equivalently, an underestimation of the out-of-band emissions. To avoid this aliasing, in this section, the time-

domain signal is oversampled by a factor $L_{\text{os}} \in \mathbb{N}$ prior to PA amplification. Oversampling is realized by zero-padding the precoded frequency-domain signal before the IDFT. Let $\boldsymbol{x}_m = [x_{m,0}, \ldots, x_{m,Q-1}]^{\mathsf{T}} \in \mathbb{C}^{Q\times 1}$ denote the precoded signal across all subcarriers at antenna m. The oversampled frequency-domain vector $\boldsymbol{x}_m^{\text{os}} \in \mathbb{C}^{Q_{\text{os}}\times 1}$ of length $Q_{\text{os}} = L_{\text{os}}Q$ is constructed as

$$[\boldsymbol{x}_m^{\text{os}}]_{q'} = \begin{cases} x_{m,q'}, & 0 \le q' < Q/2, \\ 0, & Q/2 \le q' < Q_{\text{os}} - Q/2, \\ x_{m,q'-(Q_{\text{os}}-Q)}, & Q_{\text{os}} - Q/2 \le q' < Q_{\text{os}}. \end{cases} \tag{5.34}$$

That is, the first $Q/2$ entries and the last $Q/2$ entries of $\boldsymbol{x}_m$ are preserved, while $Q_{\text{os}} - Q$ zeros are centrally inserted.

The oversampled time-domain signal is then obtained as

$$\tilde{x}_m^{\text{os}}[n] = \sqrt{L_{\text{os}}}\frac{1}{\sqrt{Q_{\text{os}}}} \sum_{q'=0}^{Q_{\text{os}}-1} [\boldsymbol{x}_m^{\text{os}}]_{q'} e^{\jmath 2\pi \frac{q'n}{Q_{\text{os}}}}, \quad n = 0, \ldots, Q_{\text{os}} - 1. \tag{5.35}$$

The scaling factor $\sqrt{L_{\text{os}}}$ compensates for the power reduction introduced by the IDFT of the zero-padded vector, as this spreads the same signal energy over L_{os} times more time-domain samples. Without the scaling factor this would reduce the average time-domain power by a factor L_{os}. With the scaling in (5.35), the average power of $\tilde{x}_m^{\text{os}}[n]$ equals the average power of the critically-sampled signal $\tilde{x}_m[n]$, which ensures that the IBO defined in Section 5.2.3 is maintained.

PA Amplification at the Oversampled Rate

The memoryless PA model of Section 5.2.3 is subsequently applied at the oversampled rate

$$\tilde{y}_m^{\text{os}}[n] = \phi_m\big(\tilde{x}_m^{\text{os}}[n]\big), \quad n = 0, \ldots, Q_{\text{os}} - 1, \tag{5.36}$$

such that the spectral regrowth produced by $\phi_m(\cdot)$ is represented over the extended bandwidth $[-L_{\text{os}}f_{\text{sym}}/2, L_{\text{os}}f_{\text{sym}}/2]$ rather than folded back into the in-band region[4].

[4] Residual aliasing from very high-order intermodulation products remains, but is negligible for sufficiently large L_{os}.

Return to the Frequency Domain

After amplification, the oversampled time-domain signal is transformed back via the DFT

$$y^{\text{os}}_{m,q'} = \frac{1}{\sqrt{L_{\text{os}}}} \frac{1}{\sqrt{Q_{\text{os}}}} \sum_{n=0}^{Q_{\text{os}}-1} \tilde{y}^{\text{os}}_{m}[n] e^{-\jmath 2\pi \frac{q'n}{Q_{\text{os}}}}, \tag{5.37}$$

where the $1/\sqrt{L_{\text{os}}}$ factor undoes the power-normalization scaling introduced in (5.35), so that the in-band subcarriers $y^{\text{os}}_{m,q'}$ share the same scaling as the original $x_{m,q}$ subcarriers. The Q in-band subcarriers are then extracted from $\boldsymbol{y}^{\text{os}}_{m}$ as

$$y_{m,q} = \begin{cases} y^{\text{os}}_{m,q}, & 0 \leq q < Q/2, \\ y^{\text{os}}_{m,q+(Q_{\text{os}}-Q)}, & Q/2 \leq q < Q. \end{cases} \tag{5.38}$$

The $Q_{\text{os}} - Q$ out-of-band bins, which now contain the spectral regrowth caused by the PA nonlinearity, are discarded. This mimics a receiver that filters the signal to the in-band bandwidth f_{sym} before demodulation, and is consistent with the Bussgang-based achievable rate analysis in Section 5.2.1, which operates on the Q in-band subcarriers only. The critically-sampled model is recovered as the special case $L_{\text{os}} = 1$, in which (5.34) is an identity and no out-of-band bins are present.

Impact on the Received Signal

Because the subsequent channel propagation $r_q = \boldsymbol{h}_q^{\mathsf{T}} \boldsymbol{y}_q + \boldsymbol{v}_q$ and the Bussgang decomposition operate on a per-subcarrier basis, the oversampling procedure only affects how $\boldsymbol{y}_q$ is generated. This implicitly assumes ideal receive filtering of the OOB content prior to demodulation. In a practical system, the channel would filter the oversampled signal across the full extended bandwidth, and the receive filter would be non-ideal. However, these effects are not considered for this analysis. For these considered assumptions, oversampling modifies the statistics of the nonlinear distortion term d_q: out-of-band intermodulation products that would otherwise alias into the in-band subcarriers are removed, leading to a more accurate characterization of the in-band SNDR. Throughout the remainder of this chapter, an oversampling factor $L_{\text{os}} = 5$ is used.

5.7.2 In-band Performance with Oversampling

First we recreate some of the results of Section 5.6, when considering oversampling. For this the GNN precoder is trained without accounting for oversampling.

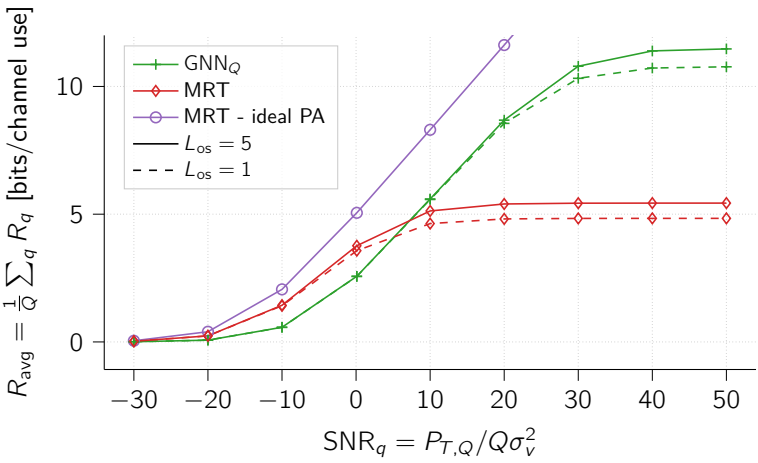


Figure 5.10: Rate vs SNR$_q$ averaged over 100 frequency flat channels ($\beta = 100$). For $M = 32, Q = 32$ and IBO $= -3$ dB. During training SNR$_q = 20$ and no oversampling is considered $L_{os} = 1$. Evaluated using no oversampling (dashed) and with oversampling $L_{os} = 5$ (full).

Consequently, the nonlinear distortion generated by the PA undergoes circular convolution in the frequency domain, which results in aliasing of OOB distortion components into the in-band subcarriers. In contrast, when oversampling is applied during evaluation, the nonlinear distortion is more accurately represented, and OOB components remain outside the signal band. One might therefore expect that a precoder trained without oversampling would learn to mitigate distortion components that are partly artificial, potentially leading to degraded performance when evaluated under a more accurate oversampled model. However, as is demonstrated next, numerical results show that although the absolute achievable rates increase when oversampling is considered (due to the absence of aliasing), the relative performance gain of the GNN precoder over conventional precoders remains largely unchanged.

In Figure 5.10 and Figure 5.11 the achievable rate for a varied SNR of GNN$_Q$ and MRT precoding is depicted for frequency flat channels and frequency selective/i.i.d. channels, both with and without considering oversampling. These figures show that considering oversampling slightly increases the achievable rate for both GNN$_Q$ and MRT as the OOB distortion is not aliased in-band, which results in an overall distortion reduction. A similar phenomenon can be seen in Figure 5.12a where the achievable rate is depicted for varying degrees of frequency selectivity (β). Also here it can be seen that the rate when considering oversampling is slightly higher

for both GNN_Q and MRT precoding.

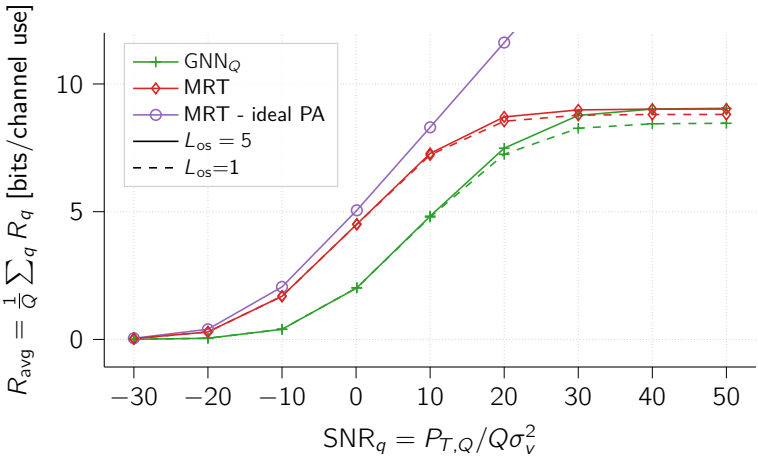
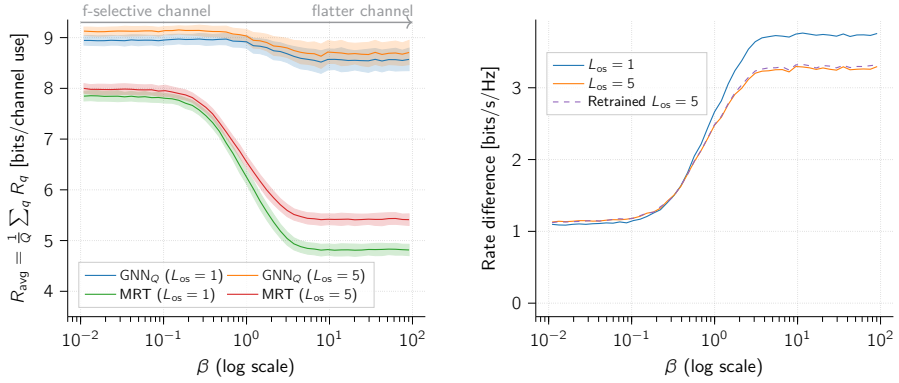


Figure 5.11: Rate vs SNR_q averaged over 100 frequency selective/i.i.d. channels. For $M = 32, Q = 32$ and IBO $= -3$ dB. During training $\text{SNR}_q = 20$ and no oversampling is considered $L_{\text{os}} = 1$. Evaluated using no oversampling (dashed) and with oversampling $L_{\text{os}} = 5$ (full).

In Figure 5.12b, the rate difference between using the GNN as compared to MRT ($R_{\text{avg}}^{\text{GNN}_Q} - R_{\text{avg}}^{\text{MRT}}$) is illustrated as a function of β, both with and without oversampling during the evaluation. In this figure, the rate difference is depicted both for GNN_Q that is trained without considering oversampling, and also for GNN_Q that is trained *with* oversampling. The figure shows that the rate difference of using GNN_Q, without considering oversampling during training, is largely maintained even when evaluated using oversampling. Only at high β values, i.e., frequency flat channels, there is a rate difference reduction of 0.45 bits/s/Hz as compared to evaluations not considering oversampling. Next to this, it is illustrated that retraining GNN_Q with oversampling yields only a negligible additional improvement over its non-oversampled counterpart. This indicate that training without oversampling is the preferred practical choice, as it captures close to all of the achievable gain while avoiding the L_{os}-fold increase in DFT/IDFT length during training.

A plausible intuitive explanation for this behavior is that the effectiveness of the GNN precoder appears to be primarily determined by the spatial structure of the nonlinear distortion, rather than by its exact spectral distribution. While oversampling changes how the distortion is distributed across frequencies, it arguably does not fundamentally alter how the distortion is generated across

antennas, how it depends on the precoded signals, or how it combines at the receiver through the channel. In particular, both the true in-band distortion (with oversampling) and the aliased in-band distortion (without oversampling) originate from the same nonlinear PA function, and may therefore share a similar spatial signature when projected onto the user channel. Given this, the GNN_Q precoder learns to steer distortion into directions that are weakly coupled to the user channel, a strategy that would remain effective regardless of whether aliasing is present. A rigorous verification of this hypothesis is left for future work.



(a) (b)

Figure 5.12: Rate vs β ± 1 standard deviation, for GNN_Q that is trained without considering oversampling (a). Rate difference between GNN_Q and MRT ($R_{\text{avg}}^{\text{GNN}_Q} - R_{\text{avg}}^{\text{MRT}}$) as a function of β, depicted for GNN_Q that is trained with and without considering oversampling (b). For $M = 32$, $Q = 32$ and IBO $= -3$ dB. During training $\text{SNR}_q = 20$. Evaluated at $\text{SNR}_q = 20$ using no oversampling and with oversampling $L_{\text{os}} = 5$.

5.7.3 Out-of-band Performance

In this subsection, we study the OOB performance for the GNN_Q and MRT precoders. While previously $Q = 32$ subcarriers were considered for computational simplicity, here the subcarrier count is increased to $Q = 128$ to improve spectral resolution. This allows for a clearer observation of spectral regrowth and OOB distortion. Consequently, GNN_Q is retrained with $Q = 128$ subcarriers and $d_h = 8Q = 1024$. Note that for this new configuration, the in-band performance in terms of R_{avg} is very closely matched to the results illustrated in Figures 5.12a

and 5.12b. This indicates that the GNN is still able to properly mitigate distortion even under these increased system settings.

Power Spectral Density

A common metric to quantify OOB distortion is the power spectral density (PSD) of the transmitted signal after the PA. The PSD is estimated using Welch's method on the oversampled time-domain signal $\tilde{y}_m^{\text{os}}[n]$ at each antenna m, such that the spectral regrowth generated by the PA nonlinearity is resolved over the full bandwidth $[-L_{\text{os}} f_{\text{sym}}/2, L_{\text{os}} f_{\text{sym}}/2]$, as discussed in Section 5.7.1. The in-band region corresponds to the normalized frequency interval $f/f_{\text{sym}} \in [-0.5, 0.5]$, while frequencies outside this interval represent OOB emissions caused by intermodulation products.

Since the PSD depends on the precoded signal at each antenna, and because the precoders GNN_Q and MRT allocate power unevenly across antennas in response to the instantaneous channel, the PSD varies from antenna to antenna. To concisely characterize this variation, Figure 5.13a and 5.13b report the best-case and worst-case PSD across the M antennas, defined as the antennas achieving the lowest and highest OOB emission power, respectively. Results are shown for a frequency-flat channel ($\beta = 100$) in Figure 5.13a and a strongly frequency-selective channel ($\beta = 10^{-3}$) in Figure 5.13b. For both channel conditions, GNN_Q exhibits higher OOB emissions as compared to MRT. This is expected, given that GNN_Q deeply saturates a subset of the antennas to mitigate the distortion in the user direction. This leads to higher worst-case OOB emissions for these saturated antennas. Conversely, the best-case GNN antenna exhibits OOB emissions comparable to MRT, confirming that only a subset of antennas is driven deep into saturation while the remaining antennas operate in a more linear regime.

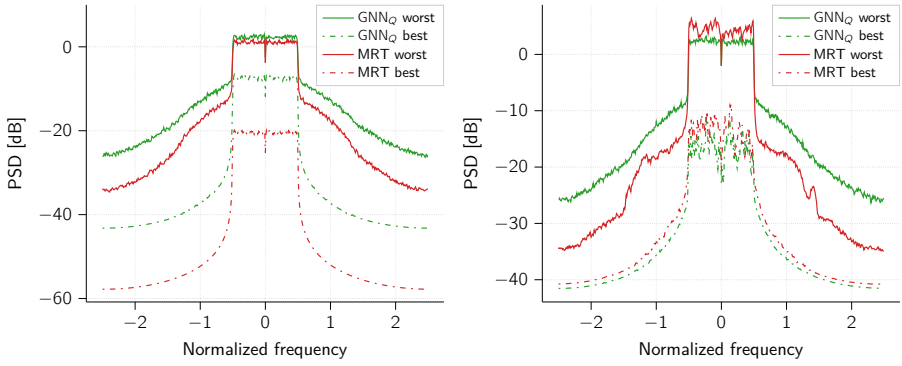


Figure 5.13: Best- and worst-case PSD across the antennas for MRT and GNN precoding for frequency-flat $\beta = 100$ (a) and frequency selective channels $\beta = 10^{-3}$ (b). With $Q = 128$, average IBO $= -3$ dB and $L_{\text{os}} = 5$.

Spatial ACLR To characterize the out-of-band emissions radiated by the antenna array in different spatial directions, we define the spatial adjacent channel leakage ratio (ACLR) as follows. Since out-of-band emissions lie, by definition, outside the nominal signal bandwidth f_{sym}, the analysis is carried out on the oversampled time-domain signals introduced in Section 5.7.1, such that the spectral regrowth generated by the PA nonlinearity is represented over the extended bandwidth $[-L_{\text{os}} f_{\text{sym}}/2, L_{\text{os}} f_{\text{sym}}/2]$. After precoding, zero-padding, and oversampled OFDM modulation, the oversampled time-domain signal at antenna m is $\tilde{y}_m^{\text{os}}[n] = \phi_m(\tilde{x}_m^{\text{os}}[n])$ for $n = 0, \ldots, Q_{\text{os}} - 1$. The far-field signal observed at angle θ' is the coherent sum

$$r_{\theta'}^{\text{os}}[n] = \boldsymbol{a}^{\mathsf{T}}(\theta')\, \tilde{\boldsymbol{y}}^{\text{os}}[n] = \sum_{m=0}^{M-1} a_m(\theta')\, \tilde{y}_m^{\text{os}}[n] \tag{5.39}$$

where $\tilde{\boldsymbol{y}}^{\text{os}}[n] = [\tilde{y}_0^{\text{os}}[n], \ldots, \tilde{y}_{M-1}^{\text{os}}[n]]^{\mathsf{T}}$ and $\boldsymbol{a}(\theta')$ is the array steering vector for a uniform linear array (ULA) in direction θ', given by

$$a_m(\theta') = e^{\jmath m \frac{2\pi}{\lambda_c} d \sin\theta'} \tag{5.40}$$

with $d = \lambda_c/2$ the antenna spacing and λ_c the carrier wavelength. The PSD $S_{r_{\theta'}^{\text{os}}}(f)$ of $r_{\theta'}^{\text{os}}[n]$ is estimated via Welch's method. Let $\mathcal{B}_{\text{in}} = [-f_{\text{sym}}/2, f_{\text{sym}}/2]$ denote the in-band frequency region, and $\mathcal{B}_{\text{adj,upper}} = [f_{\text{sym}}/2, 3f_{\text{sym}}/2]$ and $\mathcal{B}_{\text{adj,lower}} = [-3f_{\text{sym}}/2, -f_{\text{sym}}/2]$ the upper and lower adjacent out-of-band

regions [34]. The spatial ACLR at angle θ' is then defined as

$$\mathrm{ACLR}(\theta') = 10\log_{10}\left(\frac{P_{\mathrm{adj}}(\theta')}{P_{\mathrm{in}}(\theta')}\right) \tag{5.41}$$

where

$$P_{\mathrm{in}}(\theta') = \int_{f\in\mathcal{B}_{\mathrm{in}}} S_{r_{\theta'}^{\mathrm{os}}}(f)\,df \tag{5.42}$$

$$P_{\mathrm{adj}}(\theta') = \max\left\{\int_{f\in\mathcal{B}_{\mathrm{adj,upper}}} S_{r_{\theta'}^{\mathrm{os}}}(f)\,df, \int_{f\in\mathcal{B}_{\mathrm{adj,lower}}} S_{r_{\theta'}^{\mathrm{os}}}(f)\,df\right\}. \tag{5.43}$$

By evaluating $\mathrm{ACLR}(\theta')$ for all angles θ', we obtain the spatial distribution of out-of-band emissions radiated by the array.

To allow for a spatially consistent spatial ACLR plot, GNN_Q was retrained on line-of-sight (LoS) channels. A LoS channel is used because the notion of a user direction is only well-defined when the channel has a dominant angular component. The spatial ACLR evaluated on a LoS channel is depicted in Figure 5.14. From this figure, it can be seen that the GNN precoder reduces the ACLR in the user direction but, as a consequence, has an increased ACLR in the non-user directions, as compared to MRT.

These OOB results highlight a key trade-off: the in-band rate gains of GNN_Q come at the cost of increased OOB emissions. For practical deployment and regulatory compliance, this may require OOB relaxation techniques or stricter spectral mask constraints to be incorporated into the training objective. This is further discussed in Chapter 7.

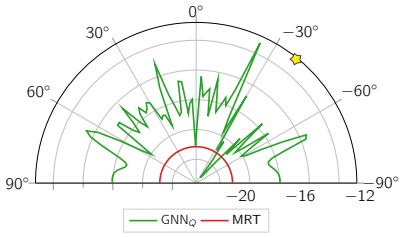


Figure 5.14: Spatial ACLR, $\mathrm{ACLR}(\theta')$ [dB] for MRT and GNN precoding for $Q = 128$, average IBO $= -3$ dB and $L_{\mathrm{os}} = 5$. Evaluated on a LoS channel, where the user is indicated by a yellow star.

5.8 Conclusion

This chapter extends distortion-aware GNN-based precoding from frequency flat conditions and single-carrier systems to multi-carrier OFDM systems for the single-user case. The performance in terms of achievable rate and the PA power consumption has been studied under a wide range of channel frequency-selectivity and PA operating points. It is shown that joint GNN precoding across subcarriers preserves the distortion-mitigation gains seen in the narrowband case when channels are flat or moderately frequency-selective, while remaining robust when channels become highly frequency-selective.

The results show that the achievable rate is maintained, and PA power consumption reductions are obtained when the distortion seen at the user combines coherently, which occurs under flat or mildly frequency-selective channels and when the PAs operate close to saturation. In this regime, GNN_Q can saturate a subset of antennas with a suitable phase shift across subcarriers to steer nonlinear distortion away from the user, achieving rate gains over MRT and PA power savings ranging from 10 to 110 watts for a fixed QoS constraint. This represents a relative PA power reduction of $34.9 - 83.5$ % as compared to MRT precoding. Conversely, when frequency selectivity becomes very strong, the distortion contributions decorrelate across subcarriers, and the spatial combining of distortion diminishes, so that the advantage of distortion-aware precoding shrinks and the performance of GNN-based precoders approaches that of MRT.

From a system design perspective, the chapter identifies operating regimes where distortion-aware GNN precoding is applicable and where it is not. It is beneficial when PAs are driven close to saturation and channels are not too frequency-selective, since then GNN_Q can simultaneously reduce PA consumption and relax linearity requirements without compromising achievable rate. In contrast, when frequency-selectivity is strong, and a stringent QoS constraint forces relatively large IBO (e.g., $R \geq 11$ bits/channel use which is equivalent to $\eta_{\text{SE}} \geq 10$ bit/s/Hz and strongly frequency selective channels), GNN-based and MRT precoders converge to similar PA power consumption. Given that GNN inference still incurs an overhead, conventional precoding (MRT) becomes preferable. The analysis of total power shows that, aside from this edge case, the DSP power required to run the GNN is small compared to the PA savings.

Finally, the impact of oversampling on both in-band and OOB performance was investigated. It is worth noting that evaluating without oversampling is slightly pessimistic in absolute terms, as the absence of OOB aliasing at $L_{\text{os}} = 5$ yields slightly higher absolute rates for both GNN_Q and MRT. Conversely, the relative

gain of GNN_Q over MRT is slightly optimistic without oversampling, as evaluating with $L_{\text{os}} = 5$ reduces the rate difference by up to 0.45 bits/s/Hz in the frequency-flat regime. Nevertheless, the observed differences are small in both absolute and relative terms, and retraining GNN_Q with oversampling yields only negligible additional improvement. This justifies the use of symbol-rate modeling ($L_{\text{os}} = 1$) during training and testing in earlier sections and chapters. With respect to OOB emission, the in-band gains of GNN_Q come at the cost of elevated OOB emissions. This is expected as a subset of the antennas is driven deep into saturation to steer distortion away from the user, resulting in higher worst-case PSD and an increased spatial ACLR in non-user directions. This trade-off between in-band performance and OOB compliance motivates the integration of spectral-mask-aware constraints into the training objective, as further discussed in Chapter 7.

The current work opens several directions for future work. First, scaling the proposed GNN systems to more transmit antennas and more sub-carriers is demanded by the current size of practical systems. Second, an extension of the analysis to multi-user systems is desirable. Here, the spatial correlation of the user channels will determine whether the distortion combines constructively or spreads out. Hence, the spatial correlation will decide whether distortion-aware precoding can still deliver reasonable gains. In particular, for scenarios where the channels are close to i.i.d. both in space and frequency, the results of this chapter suggest that limited improvements are expected. We expect that more suitable scenarios for GNN-based distortion-aware precoders are the ones that consider spatially correlated users and moderately frequency-selective channels.

Chapter 6

Learning to Quantize & Precode

This chapter is based on the following papers:

> T. Feys, L. Van der Perre, and F. Rottenberg, "Self-Supervised Learning of Graph Neural Networks for Precoding in 1-bit Quantized Large Array Systems," in *2025 IEEE International Conference on Machine Learning for Communication and Networking (ICMLCN)*, pp. 1–6, 2025 [30]
>
> T. Feys, L. Van der Perre, and F. Rottenberg, "Learning to Quantize and Precode in Massive MIMO Systems for Energy Reduction: a Graph Neural Network Approach," *IEEE Journal of Selected Topics in Signal Processing*, pp. 1–15, 2025 [31]

The previous chapters have focused on reducing the consumption of PAs. This was done by allowing the PAs to operate in an energy-efficient regime and accounting for the associated distortion through GNN-based distortion-aware precoding. In this chapter, we shift the focus to digital-to-analog converters (DACs). The main aim of this chapter is to utilize fewer bits in the DACs. By using fewer bits, the power consumption of the DACs is reduced. However, this introduces quantization-based distortion. This chapter studies the use of GNNs to perform non-linear precoding to overcome this quantization-induced distortion in DACs. In contrast to the previous chapters, non-linear precoding is utilized, where the NN learns a mapping from the channel matrix and symbol vector to the precoded quantized symbol vector. This non-linear precoder requires a new GNN architecture. Additionally, the quantization

in the DACs introduces a non-differentiable and non-continuous function. Given that the training of NNs is gradient-based, the non-differentiable DAC function is problematic. To overcome this, a straight-through Gumbel-softmax estimation of the gradient is proposed. The developed non-linear precoder addresses research question Q3 (How can NN-based nonlinear precoding architectures be designed such that the number of DAC bits can be reduced without degrading the data rate?). Next to this, the reduction in power consumption in the DACs is studied as well as the additional consumption introduced by the GNN. This is in line with research question Q4.2 (How much can the DAC consumed power be reduced, and does this outweigh the additional inference power consumption of the required NN?).

The remainder of this chapter is outlined as follows. Section 6.2 outlines the system model. In Section 6.3, the critical case in terms of distortion is identified as the few users, few bits case. Hence, the proposed precoder is developed considering few users and few bits, i.e., it does not rely on assumptions of uncorrelated distortion that are only valid when many users are present or many bits are considered, as is often done in the current literature. In Section 6.4, based on the permutation equivariance properties of the non-linear precoding problem, a GNN is proposed that captures these properties. Additionally, an unsupervised training procedure is outlined to train this GNN. This is especially challenging given the non-differentiable nature of the objective function, caused by the non-differentiable DAC functions. To overcome this, a Gumbel-softmax-based relaxation of the objective function is proposed. In Section 6.5, the complexity of the proposed GNN is computed. Section 6.6 provides extensive simulation results to validate the performance of the proposed GNN. It is shown that the proposed precoder can achieve a significant increase in terms of achievable sum rate. Radiation patterns illustrate that this gain is achieved by transmitting the distortion in non-user directions. In Section 6.7, the reduction in power consumption of the DACs is computed while taking into account the additional processing power consumption of the GNN. When solely considering the DACs consumption, it is shown that the proposed method reduces the power consumption by a factor of 4-7 and 3 for baseband and radio frequency (RF) DACs, respectively. However, when taking the GNN processing power consumption into account the power reduction is only achieved for bandwidths up to 3.5 MHz for baseband DACs, while the RF DACs can maintain a power reduction of 2.9 for higher bandwidths. Notably, indirect effects, which further reduce the power consumption, such as reduced fronthaul consumption and reductions in other components, are not considered in this analysis. Finally, Section 6.9 concludes the chapter.

6.1 Introduction

6.1.1 Problem Statement

In a fully digital massive MIMO system, each radio frequency chain is typically equipped with its own, analog-to-digital converter (ADC), DAC, PA and antenna, among others. The use of many radio frequency chains increases both the hardware complexity and the power consumption. To combat this, in this chapter we focus on reducing the bit-width of DACs. This is motivated by the current move towards higher carrier frequencies and higher bandwidths [15], which further increases the complexity and the power consumption of the DACs. The reason for this increased power consumption is threefold. First, the power consumption of typical current steering DACs scales exponentially with the number of bits and linearly with the sampling frequency [13]. This raises concerns given the ever-increasing bandwidth of wireless systems. Second, radio frequency digital-to-analog converters (RF-DACs) are becoming increasingly popular in wireless systems due to their ability to directly synthesize the RF signal at the desired carrier frequency, which reduces analog complexity [87]. This is possible by relying on a high oversampling factor to place the signal at the desired carrier frequency, bypassing traditional analog upconversion. However, due to the increased oversampling factor, coupled with the current move towards higher carrier frequencies and higher bandwidths, the power consumption of these RF-DACs is significantly higher as compared to traditional DACs. To offset this increased power consumption, we investigate how the bit-width can be reduced, providing reduced power consumption in the DACs, while maintaining system performance. Finally, the spurious free dynamic range, that is the frequency range over which the DAC operates free from unwanted spurious signals, is limited by the resolution of the DAC [88]. Hence, high resolution DACs become problematic at higher carrier frequencies and bandwidths, as additional nonlinearities are introduced. Next to this, decreasing the bit-width can have significant indirect benefits. For instance, the power consumption of the digital fronthaul, which connects the baseband unit with the antenna array, can be significantly reduced when fewer bits are considered. Alternatively, if the bit-width reduction can be propagated into the baseband processing this would have a big impact on its power consumption.

Consequently, low-resolution DACs are vital to overcome these hardware limitations and reduce the power consumed in the many DACs. However, by using lower resolution DACs a significant quantization error is introduced, which limits the signal-to-distortion ratio (SDR) and the achievable data rates. To overcome this, we present an approach to limit this quantization distortion by leveraging the

degrees of freedom available from the many antennas at the BS. This is done by proposing a NN-based non-linear precoding scheme that mitigates this distortion. By doing so, the DACs can utilize fewer bits, while a high achievable data rate can be maintained in the network.

6.1.2 State-of-the-Art Quantization in Massive MIMO

Quantization in Massive MIMO systems has been widely studied. A rich body of literature focuses on the effects of ADCs at the receiver [89, 90, 91, 92, 93]. In this work, we focus on the effects of DAC distortion. Two trends have emerged when dealing with DAC distortion in MIMO systems. On the one hand, there is a focus on 1-bit systems, largely driven by their analytical tractability and simplified hardware design. On the other hand, several works tackle the more general few-bit case. However, due to the intractability, simplifying assumptions have been made which are only valid when many users or many bits are considered. As shown in this work, distortion due to the DACs is most limiting when few users are present. This is because distortion becomes more spatially spread out as the number of users rises, similar to the case of non-linear PA distortion, as seen in the previous chapters [34]. Hence, there is a need for a solution to the quantization-aware precoding problem for the case of few users and few bits.

1-bit DACs

In [94], a 1-bit quantized ZF precoder is studied. It is shown that when the number of transmit antennas goes to infinity, the received signals are scaled versions of the desired symbols. Next to this, in [95] a two-stage digital-analog precoder is proposed that takes into account one-bit DACs at the transmitter and one-bit ADC at the receiver. Additionally, in [96, 97] several one-bit non-linear precoders are proposed i) based on a semidefinite relaxation leading to a convex problem, ii) based on a squared l_∞-norm relaxation which also leads to a convex problem and iii) based on sphere decoding rather than solving the non-linear quantized precoding problem, which essentially is an NP-hard closest-vector problem, through exhaustive search.

Few-bit DACs

In [98], minimum mean squared error (MMSE) linear processing is proposed where 4-6 bit DACs are deployed at the BS. However, this work relies on the high-

resolution assumption to specify the quantization error, i.e., when the number of bits b is large, the quantization error of an optimal non-uniform quantizer approximately scales with 2^{-2b}. The work in [99] optimizes the energy efficiency while taking into account the power consumption and distortion due to low-resolution ADCs and DACs. However, the distortion covariance matrix is assumed to be diagonal (i.e., uncorrelated distortion). This assumption holds when many users with an i.i.d. Rayleigh fading channel are considered [100]. Hence, this assumption does not hold in the most critical case for quantization distortion, i.e., the few user-case. Similarly, in [101] the quantized single-user MIMO case is studied. The energy efficiency of digital and hybrid precoders is evaluated. However, this study also relies on the assumption of uncorrelated distortion, which only holds when many users are present.

In summary, there is a need for transmit processing that addresses the case of few-bit DACs when few users are present. In this study, we overcome this gap by proposing neural network-based non-linear precoders that are trained for both the one-bit and few-bit DAC cases, without relying on assumptions that only hold for high-resolution DACs or a high number of users.

6.1.3 State-of-the-Art Neural Network-based Precoding

NNs have become an interesting approach for learning to precode. Some works focus on learning NN-based precoders as a low-complexity alternative to high-complexity conventional precoding algorithms [102, 103, 104]. Furthermore, NN-based precoding has been used to solve precoding problems that do not have a simple solution and where conventional algorithms show limited performance [103, 28, 27]. In these works a wide range of NN architectures is used ranging from multilayer perceptrons (MLPs), convolutional neural networks (CNNs), transformers and GNNs. Each of these architectures has a certain structure that limits the types of functions these NNs can learn. As explained in Chapter 3 the aim of selecting a certain architecture is to reduce the hypothesis space of the NN in an informed manner such that the desired function is still covered by this reduced hypothesis space. Under these conditions the reduced expressivity of the model does not affect the learning performance [65, 28]. Note that this is closely linked to the bias-variance trade-off [105]. A model with a large hypothesis space will have a low bias but high variance, while reducing the hypothesis space leads to lower variance, but might increase the bias. The goal is to find a hypothesis space with an optimal trade-off such that both the bias and variance are minimized.

As shown in Chapter 3 and other recent studies [64, 67], GNNs are especially well-suited for linear precoding. This is due to their permutation equivariance properties. Simply put, if the order of the rows/columns in the input matrix to the NN is permuted, the order of the rows/columns in the output matrix should be permuted accordingly. Concerning the linear precoding problem, if the rows/columns of the channel matrix (input) are permuted, the rows/columns of the precoding matrix (output) should be permuted accordingly. This is an inherent property of linear precoders, which is respected by using GNNs. Hence, by using GNNs the hypothesis space is reduced to only include permutation equivariant functions, thereby reducing the variance while keeping the bias low.

Concerning quantization-aware precoding, several works have introduced the use of NNs for the 1-bit case. In [106], the authors propose to unfold the biconvex 1-bit precoding algorithm from [107]. Next to this, in [108] non-linear precoding for the 1-bit case is learned in a supervised manner.

More broadly, the quantized precoding problem can be seen as a combinatorial problem, as the goal is to select the best output level at each DAC from a finite set. Our work is inspired by a sub-field of deep learning, known as neural combinatorial optimization (NCO) [109, 110, 111]. NCO aims at training NNs that can solve combinatorial problems. Given the NP-hard nature of many combinatorial problems, it is often hard to obtain optimal solutions that can be used as labels. Hence, supervised learning for combinatorial optimization is often costly or impossible. Consequently, NCO is typically done in combination with reinforcement learning, where the proposed solution by the NN is evaluated. This evaluation is then used as a reward signal to update the weights of the NN [109]. More recently, one-shot neural combinatorial solvers have been introduced that do not rely on reinforcement learning [112]. These methods directly predict the decision variables for the problem. The decision variables are then evaluated using a differentiable objective function. This differentiable objective function is then used as a self-supervised loss function, to update the weights of the NN through stochastic gradient descent [112]. The benefit of this methodology is *i)* that no labels are required, *ii)* unsupervised learning typically generalizes better as compared to the mimic learning that happens in the supervised setting, which can be attributed to the unsupervised loss function capturing more underlying physical insights of the problem, and *iii)* it is empirically more efficient than reinforcement learning [112]. The main drawback of this method is that the objective function should be differentiable, which is not always the case. In this work, we follow the one-shot neural combinatorial solver methodology. Given the non-differentiable nature of our objective function (caused by the non-differentiable DAC functions), we utilize a Gumbel-softmax-based [113] relaxation which renders the objective function differentiable in the backward pass but still discrete in the forward pass.

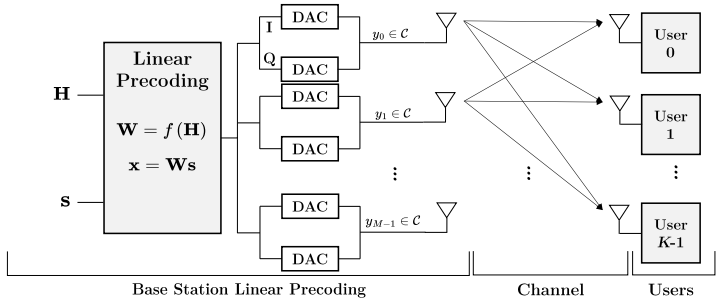


Figure 6.1: Simplified system overview of linear quantized precoding.

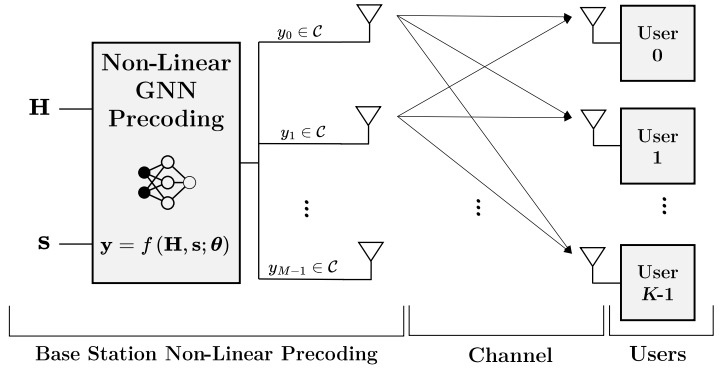


Figure 6.2: Simplified system overview of non-linear quantized precoding.

6.2 System Model

6.2.1 Linear Quantized Precoding

An overview of the linear quantized precoding system model is given in Figure 6.1. In this case we consider a BS with M transmit antennas and K single antenna users, operating in the downlink. When considering digital full-precision linear precoding, the precoded vector is obtained as $\boldsymbol{x} = \boldsymbol{W}\boldsymbol{s}$, where $\boldsymbol{x} \in \mathbb{C}^M$ are the precoded symbols, $\boldsymbol{W} \in \mathbb{C}^{M \times K}$ is the precoding matrix and $\boldsymbol{s} \in \mathbb{C}^K$ is the symbol vector, where the symbols between different users are uncorrelated and zero mean. When a single user is present we consider MRT precoding, with precoding matrix

$\boldsymbol{W}^{\text{MRT}} = \alpha_{\text{norm}}\boldsymbol{H}^*$. When multiple users are present ZF precoding is considered as $\boldsymbol{W}^{\text{ZF}} = \alpha_{\text{norm}}\boldsymbol{H}^*(\boldsymbol{H}^{\intercal}\boldsymbol{H}^*)^{-1}$. Both for MRT and ZF precoding the power normalization constant is defined as $\alpha_{\text{norm}} = \sqrt{P_T/\text{Tr}(\boldsymbol{W}\boldsymbol{W}^H)}$ so that the total transmit power is normalized to P_T. After precoding, the digital baseband signal $\boldsymbol{x}$ is quantized. This is done for the real and imaginary dimensions per antenna

$$\begin{aligned}\boldsymbol{y} = \mathcal{Q}(\boldsymbol{x}) = [\mathcal{Q}(\Re\{x_0\}) + \jmath\mathcal{Q}(\Im\{x_0\}), \cdots, \\ \mathcal{Q}(\Re\{x_{M-1}\}) + \jmath\mathcal{Q}(\Im\{x_{M-1}\})]^{\intercal}\end{aligned} \tag{6.1}$$

where $\mathcal{Q}(\cdot) : \mathbb{C}^M \mapsto \mathcal{C}^M$ denotes the quantization function which characterizes the combined operation of the $2M$ DACs at the BS and $\mathcal{Q}(\cdot) : \mathbb{R} \mapsto \mathcal{L}$ denotes the real-valued quantizer. These quantizers are further detailed in Section 6.2.2.

This quantization can be characterized by the additive quantization noise model (AQNM), originally proposed by [114], which is a special case of the Bussgang decomposition with the nonlinear function being a quantizer that represents each quantization interval by its mean value [43]. We follow the convention of the Bussgang decomposition

$$\boldsymbol{y} = \mathcal{Q}(\boldsymbol{x}) = \boldsymbol{\Phi}_\alpha\boldsymbol{x} + \boldsymbol{q} \tag{6.2}$$

where the distortion $\boldsymbol{q}$ is uncorrelated with $\boldsymbol{x}$, i.e., $\mathbb{E}\left(\boldsymbol{x}\boldsymbol{q}^H\right) = \boldsymbol{0}$. The Bussgang matrix is given by $\boldsymbol{\Phi}_\alpha = \text{diag}\left(\alpha_0, \cdots, \alpha_{M-1}\right) \in \mathbb{R}^{M\times M}$, where $\alpha_m = \frac{\mathbb{E}(y_m x_m^*)}{\mathbb{E}(|x_m|^2)}$. The link with the AQNM can be made by defining the normalized mean squared quantization error (NMSQE) after the m^{th} DAC as

$$\beta_m = \frac{\mathbb{E}\left((x_m - y_m)(x_m - y_m)^*\right)}{\mathbb{E}\left(|x_m|^2\right)} \tag{6.3}$$

and relating the Bussgang gain to the NMSQE after each DAC as $\boldsymbol{\Phi}_\alpha = \boldsymbol{I}_M - \boldsymbol{\Phi}_\beta$, where $\boldsymbol{\Phi}_\beta = \text{diag}\left(\beta_0, \cdots, \beta_{M-1}\right) \in \mathbb{R}^{M\times M}$.

6.2.2 Quantization

Quantization of a Complex-Valued Vector

The precoded (unquantized) vector is given by $\boldsymbol{x} \in \mathbb{C}^M$. After quantization, the precoded vector is given by $\boldsymbol{y} \in \mathcal{C}^M$, where the set $\mathcal{C}$ is the set of complex output levels/codebook. When considering infinite precision the set $\mathcal{C}$ coincides with $\mathbb{C}$. However, due to the use of DACs, the set $\mathcal{C}$ has a finite cardinality. The set of real-valued DAC outputs is given by $\mathcal{L} = \{l_0, \cdots l_{L-1}\} \subset \mathbb{R}$, where l_i are

the output values/levels and $L = |\mathcal{L}| = 2^b$ is the number of quantization levels with b the number of considered bits per real dimension. For each BS antenna the same output set is considered for the real and imaginary parts. The set of complex-valued DAC outputs is then given by $\mathcal{C} = \mathcal{L} \times \mathcal{L}$. This can be seen as applying one real-valued DAC to the real part of the precoded symbol at each antenna and one real-valued DAC to the imaginary part. The real-valued quantizer partitions the real line $\mathbb{R}$ into L cells $\mathcal{R}_i$ for $i = 0, 1, \cdots, L-1$. The ith cell is defined as $\mathcal{R}_i = \{x \in \mathbb{R} : \mathcal{Q}(x) = l_i\}$ [115]. From this definition, it follows that $\bigcup_i \mathcal{R}_i = \mathbb{R}$ and $\mathcal{R}_i \cap \mathcal{R}_j = \emptyset$. Additionally, each cell is fully described by its boundary points/thresholds $\mathcal{T} = \tau_0, \cdots, \tau_L$, $\mathcal{R}_i = (\tau_i, \tau_{i+1}]$ for $i = 0, \cdots, L-1$ [115]. For quantizers operating on unbounded inputs typically $\tau_0 = -\infty$ and $\tau_L = +\infty$. The optimal values of the output labels and thresholds depend on the type of quantization used and the input distribution, as described in the next section.

Non-Uniform Quantization

The input distribution in this work is Gaussian, which is a valid assumption for OFDM systems [56]. For a real-valued Gaussian input, uniform quantization is not optimal in the minimum mean squared quantization error (MMSQE) sense [115]. A MMSQE non-uniform quantizer can be found by solving $\min_{\mathcal{L},\mathcal{T}} \mathbb{E}\left((x - \mathcal{Q}(x))^2\right) = \sum_{i=1}^{N} \int_{\mathcal{R}_i} (x - l_i)^2 f_X(x) dx$. The Max-Lloyd algorithm as described in [115, 116] can be used to find the optimal output levels and thresholds. This is a hill-climbing algorithm, which can lead to local optima in the solution, hence it is advised to run the algorithm for several initialization points. This is done to obtain the output levels and thresholds used in this work.

Input Normalization

To achieve optimal quantization the input distribution to the DAC should be $\mathcal{CN}(0, 1)$. The transmit symbols are $s_k \sim \mathcal{CN}(0, 1)$. However, the input to the DAC is not the transmit symbol but the precoded symbol $x_m = \sum_k w_{m,k} s_k = \boldsymbol{w}_m^{\mathsf{T}} \boldsymbol{s}$. The precoded symbol is a scaled random variable that is complex normal distributed. Its power (variance) can be computed as $\mathbb{E}\left(|x_m|^2\right) = \|\boldsymbol{w}_m\|_2^2$. To adjust the power of the input signal (per antenna chain) to the dynamic range of the DAC, an input normalization[1] scales the input as $\tilde{x}_m = \frac{1}{\sqrt{\rho_m}} x_m$, where ρ_m is a scalar normalization factor. If $\rho_m = \|\boldsymbol{w}_m\|_2^2$ the signal power is rescaled to one. This normalized signal is sent to the DAC to be quantized as $\tilde{y}_m = \mathcal{Q}(\tilde{x}_m)$.

[1] This is similar to the automatic gain control typically applied at the receiver before the ADC.

After quantizing this normalized signal, the output of the DAC should be denormalized as $y_m = \sqrt{\rho_m}\tilde{y}_m$. This denormalization can be taken into account by the amplification stage, while the normalization can be performed in the digital signal processing.

6.2.3 Non-Linear Neural Network-based Quantized Precoding

NN-based non-linear precoding is considered, as depicted in Figure 6.2. This entails that a NN is used to learn a mapping from the channel matrix and symbol vector to the precoded quantized symbol vector. The function we want to learn can be represented as

$$\boldsymbol{y}_{\text{NL}} = f\left(\boldsymbol{H}, \boldsymbol{s}; \boldsymbol{\theta}\right) \tag{6.4}$$

where $f(\cdot,\cdot) : \mathbb{C}^{M\times K} \times \mathbb{C}^K \mapsto \mathcal{C}^M$ represents the NN, i.e., a learned non-linear function mapping parametrized by its trainable weights $\boldsymbol{\theta}$. Given that there is no explicit quantization step for non-linear precoding, the Bussgang decomposition with respect to $\boldsymbol{x}$ can no longer be computed. More generally, we define the Bussgang decomposition between the precoded quantized signal $\boldsymbol{y}$ and the intended transmit symbols $\boldsymbol{s}$ as

$$\boldsymbol{y} = \boldsymbol{G}\boldsymbol{s} + \boldsymbol{q} \tag{6.5}$$

where the Bussgang gain matrix is no longer a square diagonal matrix, it is now defined as $\boldsymbol{G} = \mathbb{E}\left(\boldsymbol{y}\boldsymbol{s}^H\right)\left(\mathbb{E}\left(\boldsymbol{s}\boldsymbol{s}^H\right)\right)^{-1} = \mathbb{E}\left(\boldsymbol{y}\boldsymbol{s}^H\right)$. The link with the Bussgang decomposition under linear precoding in (6.2) is easily made using $\boldsymbol{x} = \boldsymbol{W}\boldsymbol{s}$, as $\boldsymbol{G} = \boldsymbol{\Phi}_\alpha \boldsymbol{W}$.

6.2.4 Achievable Sum Rate

Regardless of which precoding technique is considered, the received signal can be written as

$$\boldsymbol{r} = \boldsymbol{H}^{\mathsf{T}}\boldsymbol{y} + \boldsymbol{v} \tag{6.6}$$

where $\boldsymbol{r} \in \mathbb{C}^K$ is the received signal vector, $\boldsymbol{H} \in \mathbb{C}^{M\times K}$ the channel matrix, $\boldsymbol{y} \in \mathcal{C}^M$ the precoded and quantized transmit symbols and $\boldsymbol{v} \in \mathbb{C}^K$ additive white Gaussian noise with covariance matrix $\sigma_v^2 \boldsymbol{I}_K$. Using (6.5), the received signal for user k can

be written as

$$r_k = \underbrace{\boldsymbol{h}_k^{\mathsf{T}} \boldsymbol{g}_k s_k}_{\text{intended signal}} + \underbrace{\sum_{k' \neq k} \boldsymbol{h}_k^{\mathsf{T}} \boldsymbol{g}_{k'} s_{k'}}_{\text{user interference}} + \underbrace{\boldsymbol{h}_k^{\mathsf{T}} \boldsymbol{q}}_{\text{distortion}} + \underbrace{v_k}_{\text{noise}}. \tag{6.7}$$

Hence the signal-to-noise-and-interference-and-distortion ratio (SNIDR) for user k can be written as

$$\text{SNIDR}_k = \frac{|\boldsymbol{h}_k^{\mathsf{T}} \boldsymbol{g}_k|^2}{\sum_{k' \neq k} |\boldsymbol{h}_k^{\mathsf{T}} \boldsymbol{g}_{k'}|^2 + \boldsymbol{h}_k^{\mathsf{T}} \mathbb{E}\left(\boldsymbol{q}\boldsymbol{q}^H\right) \boldsymbol{h}_k^* + \sigma_v^2}. \tag{6.8}$$

An achievable sum rate R_{sum}, i.e., a lower bound on the capacity, can be obtained by considering a worst-case scenario, where the noise and quantization distortion are jointly Gaussian distributed and independent from the data symbols

$$R_{\text{sum}} = \sum_{k=0}^{K-1} \log_2\left(1 + \text{SNIDR}_k\right). \tag{6.9}$$

6.2.5 Radiation Patterns

In Sections 6.3 and 6.6 a number of radiation patterns are evaluated. For this, a pure LoS channel $\boldsymbol{h}_{\text{los}}(\phi_k)$ is considered, the channel coefficient from antenna m to user k is

$$h_{m,k} = e^{-\jmath m \frac{2\pi}{\lambda_c} d \cos(\phi_k)}. \tag{6.10}$$

Here $m\frac{2\pi}{\lambda_c} d \cos(\phi_k)$ is the antenna-dependent phase shift when considering a ULA and a narrowband system, λ_c is the carrier frequency, $d = \lambda_c/2$ the antenna spacing and ϕ_k the user angle. The radiation pattern of the intended signal in an arbitrary direction $\tilde{\phi}$ can be written as

$$P_{\text{lin}}(\tilde{\phi}) = \mathbb{E}\left(|\boldsymbol{h}_{\text{los}}^{\mathsf{T}}(\tilde{\phi}) \boldsymbol{G}\boldsymbol{s}|^2\right) = \boldsymbol{h}_{\text{los}}^{\mathsf{T}}(\tilde{\phi}) \boldsymbol{G}\boldsymbol{G}^H \boldsymbol{h}_{\text{los}}^*(\tilde{\phi}). \tag{6.11}$$

The radiation pattern of the quantization distortion is given by

$$P_{\text{dist}}(\tilde{\phi}) = \mathbb{E}\left(|\boldsymbol{h}_{\text{los}}^{\mathsf{T}}(\tilde{\phi}) \boldsymbol{q}|^2\right) = \boldsymbol{h}_{\text{los}}^{\mathsf{T}}(\tilde{\phi}) \mathbb{E}\left(\boldsymbol{q}\boldsymbol{q}^H\right) \boldsymbol{h}_{\text{los}}^*(\tilde{\phi}). \tag{6.12}$$

The SDR radiation pattern is defined as $P_{\text{SDR}}(\tilde{\phi}) = \frac{P_{\text{lin}}(\tilde{\phi})}{P_{\text{dist}}(\tilde{\phi})}$.

6.3 Spatial Characteristics of Quantization-Induced Distortion

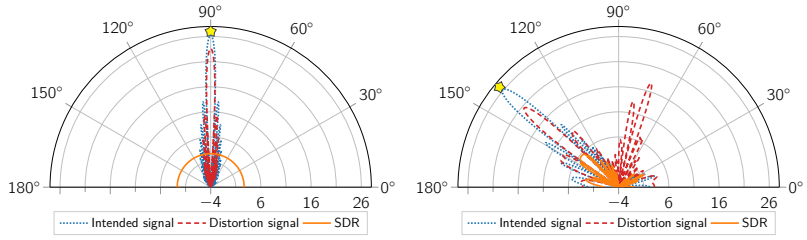
When considering linear precoding techniques such as MRT and ZF, followed by a quantization step, the radiation patterns of both the linear and distortion signals can be characterized according to (6.11) and (6.12). In this section, these numerically obtained radiation patterns are analyzed in order to identify scenarios where the quantization distortion is most critical.

In Figure 6.3 the radiation pattern of the distortion, the intended signal and SDR are depicted for $K = 1$, $b = 1$ and $M = 32$. For some pathological cases (for instance $\phi \in \{0°, 60°, 90°, 120°, 180°\}$), the beamforming gain assigned to the intended and distortion signal are identical, which leads to an SDR that is uniformly distributed in space, as seen in Figure 6.3a. In most other cases, the distortion receives a beamforming gain in multiple directions, with the user direction being the dominant one, as seen in Figure 6.3b. This leads to a small SDR in the user direction. Increasing the number of bits generally leads to identical but scaled radiation patterns, where the intended signal increases, the distortion decreases and the SDR increases.

In Figure 6.4 the radiation pattern of the distortion, the intended signal and SDR are depicted for the multi-user case. In Figure 6.4a two users are present, in this case, the distortion is radiated in a number of directions, with the two user directions being the most dominant ones. When the number of users increases to six in Figure 6.4b the distortion is nearly uniformly distributed, hence receiving little to no beamforming gain. From these figures, it is clear that the distortion radiation pattern is more spatially spread out the more users are present. This is similar to the case of non-linear PA distortion [34]. Given this spatially spread out distortion, the SDR radiation pattern shows a clear peak in the user directions. Hence the more users are present, the less they are affected by distortion. Note that this discussion purely focuses on the distortion, while the low resolution of DACs can also lead to additional inter-user interference given that the low resolution leads to poorer interference cancellation. Nevertheless, when purely focusing on the distortion the most critical case arises when few users are present.

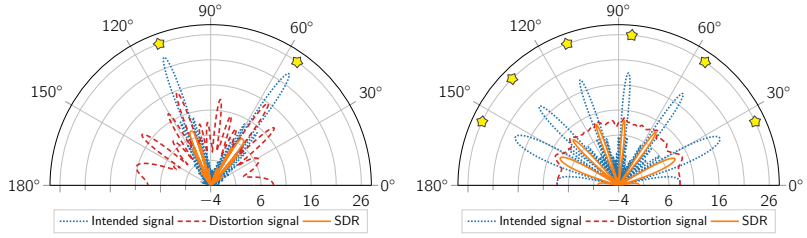
6.4 Neural Networks for Precoding

In this work, we consider NN-based non-linear precoding as described by (6.4). In this section, the training procedure is outlined, followed by the NN architecture



(a) MRT $\phi = 90°$ (b) MRT $\phi = 140°$

Figure 6.3: Radiation pattern of the intended signal $P_{\text{lin}}(\tilde{\phi})$, distortion signal $P_{\text{dist}}(\tilde{\phi})$ and SDR $P_{\text{SDR}}(\tilde{\phi})$ in [dB] for a 1-bit DAC $b = 1$, $M = 32$, $K = 1$ for MRT precoding for different user angles. Users are indicated by a star.



(a) ZF $M = 32$, $K = 2$, $b = 1$ (b) ZF $M = 32$, $K = 6$, $b = 1$

Figure 6.4: Radiation pattern for ZF of the intended signal $P_{\text{lin}}(\tilde{\phi})$, distortion $P_{\text{dist}}(\tilde{\phi})$ and SDR $P_{\text{SDR}}(\tilde{\phi})$ in [dB] for a 1-bit DAC. Users are indicated by a star, the user angles are $55°, 110°$ in (a) and $\{25°, 55°, 85°, 110°, 135°, 155°\}$ in (b).

that is selected.

6.4.1 General Training Procedure

The optimization problem we aim to solve in order to perform quantized non-linear precoding can be formulated as

$$\max_{\boldsymbol{y}_{\text{NL}} \in \mathcal{C}^M} \quad R_{\text{sum}}\left(\boldsymbol{y}_{\text{NL}}\right) \quad \text{s.t.} \quad \mathbb{E}\left(\|\boldsymbol{y}_{\text{NL}}\|_2^2\right) \leq P_T \tag{6.13}$$

where the sum rate is computed according to (6.9). During training, the NN outputs a quantized precoded vector $\boldsymbol{y}_{\text{NL}}$. This precoded quantized vector can be used to compute the loss function, which maximizes the expected sum rate over the channel and symbol distributions $p(\boldsymbol{H})$, $p(\boldsymbol{s})$, to obtain the neural network

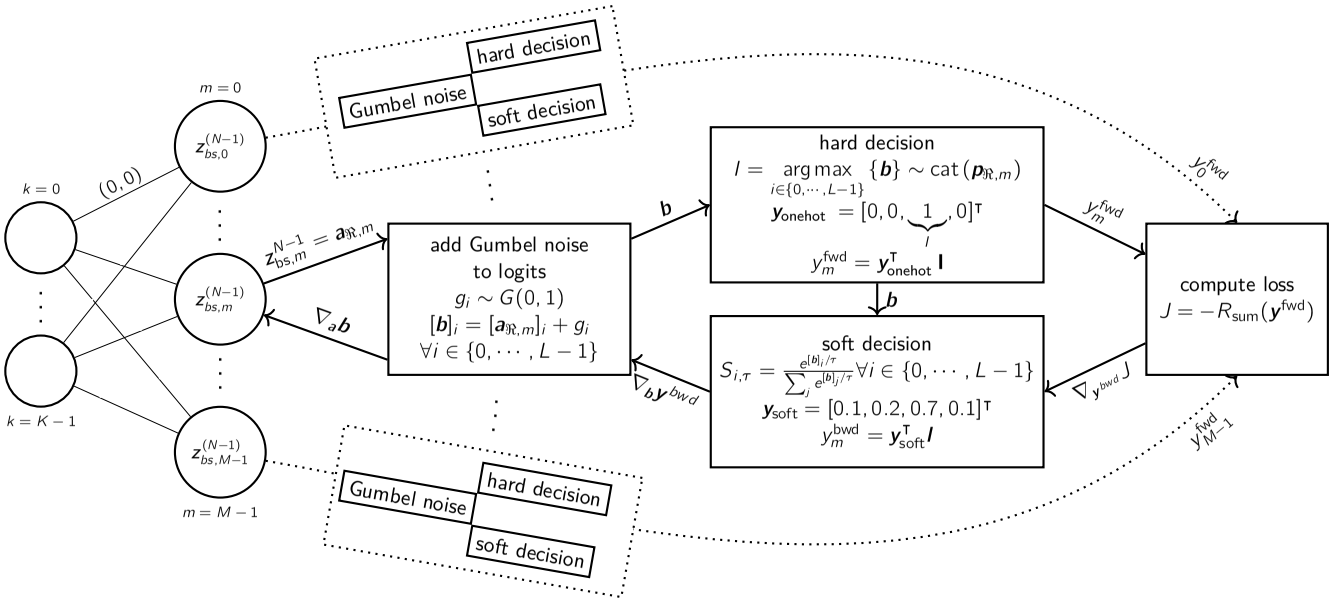


Figure 6.5: Overview of the forward and backward pass during training. For clarity, only the real part is considered in this figure but the same applies for the imaginary part $a_{\Im,m}$. The normalization step is omitted for simplicity. The vector I contains all quantization levels, hence $y_{\text{onehot}}^\intercal I$ selects the desired output level. In this example, $b = 2$ bits are considered, hence $L = 4$.

parameters

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \ \mathbb{E}_{\boldsymbol{H}\sim p(\boldsymbol{H}),\boldsymbol{s}\sim p(\boldsymbol{s})}[-R_{\text{sum}}\left(f\left(\boldsymbol{H},\boldsymbol{s};\boldsymbol{\theta}\right)\right)]\,. \tag{6.14}$$

In practice, this expectation is intractable and is approximated by the empirical average over a training set $\mathcal{D} = \{\boldsymbol{H}^{(i)}, \boldsymbol{s}^{(i,j)}\}_{i=0,j=0}^{N_{\text{train}}-1,N_s-1}$ containing N_{train} channel realizations and, for each channel, N_s independently drawn symbol vectors. This results in the empirical risk minimization problem

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \ -\frac{1}{N_{\text{train}}N_s}\sum_{i=0}^{N_{\text{train}}-1}\sum_{j=0}^{N_s-1} R_{\text{sum}}(f(\boldsymbol{H}^{(i)}, \boldsymbol{s}^{(i,j)};\boldsymbol{\theta})). \tag{6.15}$$

Here the sum rate R_{sum} is given by (6.9). Given that the value of this sum rate directly depends on the output of the NN (i.e., the precoded quantized vector), the gradients of this loss function, with respect to the weights of the NN $\boldsymbol{\theta}$ can directly be computed using backpropagation. Given these gradients, the parameters of the NN are updated using the Adam optimizer [61], on mini-batches drawn from $\mathcal{D}$. Note that during training, for each channel realization, the NN is run N_s times, i.e., for each symbol. The average power is then normalized to P_T by taking the expectation over these N_s symbols according to

$$\alpha = \frac{\sqrt{P_T}}{\mathbb{E}\left(\|\boldsymbol{y}_{\text{NL}}\|_2^2\right)}, \quad \boldsymbol{y}_{\text{norm}} = \alpha\boldsymbol{y}_{\text{NL}}. \tag{6.16}$$

Given the discrete nature of quantization, obtaining a gradient (or estimate thereof) of the loss function with respect to the weights of the NN is non-trivial. In the next subsection, the link between quantized precoding and classification is made, followed by a method to obtain a gradient estimate, which can be used to train the NN in an unsupervised manner.

6.4.2 Unsupervised Learning of Non-Linear Precoding

Given the discrete nature of the quantized precoding problem, the selection of the appropriate output levels can be seen as a classification problem per antenna, where the appropriate output levels need to be selected. However, the optimal output levels are unknown, hence there are no labels available to train the NN in a supervised manner.

Intermediate Probability Mapping

As a solution, let's consider unsupervised learning of non-linear precoding. For this, the NN uses an intermediate mapping to probability vectors, rather than

producing the correct output levels directly. The output of the NN represents a probability vector per real and imaginary part of each antenna

$$\boldsymbol{p} = f(\boldsymbol{H}, \boldsymbol{s}; \boldsymbol{\theta}), \quad \text{with} \quad \boldsymbol{p} \mid \boldsymbol{H}, \boldsymbol{s} \tag{6.17}$$

$$= [\boldsymbol{p}_{\Re,0}, \boldsymbol{p}_{\Im,0}, \cdots, \boldsymbol{p}_{\Re,M-1}, \boldsymbol{p}_{\Im,M-1}]^{\mathsf{T}} \tag{6.18}$$

where the intermediate mapping is $f(\cdot,\cdot) : \mathbb{C}^{M\times K} \times \mathbb{C}^{K} \mapsto [0,1]^{2M|\mathcal{L}|}$ and $\boldsymbol{p} \in [0,1]^{2M|\mathcal{L}|}$ represents the concatenation of the probability vectors for the real and imaginary parts of each antenna, where $\boldsymbol{p}_{\Re,m}, \boldsymbol{p}_{\Im,m} \in [0,1]^{|\mathcal{L}|}$ represent the probability vectors for the real and imaginary parts of the DAC output at antenna m. These probability vectors are conditioned on the channel matrix and symbol vector. However, this conditioning is often omitted for notational convenience. At inference time, the true output levels of DAC m can be found as the output levels $l_i, l_j \in \mathcal{L}$ that are assigned the highest probability value. This gives $y_m = l_i + l_j\jmath$, with

$$i = \underset{i\in\{0,\cdots,L-1\}}{\arg\max}\ \boldsymbol{p}_{\Re,m}, \quad j = \underset{j\in\{0,\cdots,L-1\}}{\arg\max}\ \boldsymbol{p}_{\Im,m}. \tag{6.19}$$

Each of the obtained probability vectors is a result of a softmax activation. The elements of these vectors can be interpreted as the parameters of a categorical distribution conditioned on $\boldsymbol{H}$ and $\boldsymbol{s}$, where each category (output level in our case) is assigned a certain probability of being the correct output level. The logits/the raw outputs of the NN (i.e., the output before applying the softmax activation) for the real and imaginary parts of antenna m are defined as $\boldsymbol{a}_{\Re,m}$, $\boldsymbol{a}_{\Im,m}$ respectively. The softmax of vector $\boldsymbol{a}_{\Re,m}$ is computed as[2]

$$[\boldsymbol{p}_{\Re,m}]_i = [\text{softmax}(\boldsymbol{a}_{\Re,m})]_i = \frac{e^{[\boldsymbol{a}_{\Re,m}]_i}}{\sum_j e^{[\boldsymbol{a}_{\Re,m}]_j}} \tag{6.20}$$

where $[\text{softmax}(\boldsymbol{a}_{\Re,m})]_i$ is the i-th element of the softmax-transformed vector $\boldsymbol{p}_{\Re,m}$. This scales all values of $\boldsymbol{a}_{\Re,m}$ between zero and one and ensures that the summation of all elements of the vector equals one, hence the output values of the softmax function can be interpreted as probabilities.

Overcoming Non-Differentiability

If we know the correct labels for $\boldsymbol{p}_{\Re,m}, \boldsymbol{p}_{\Im,m}$ we can learn in a supervised manner by minimizing the cross entropy loss between the output of the softmax layers and

[2]For the remainder of this discussion we will only refer to the real part of antenna m, $\boldsymbol{a}_{\Re,m}$, however the same applies to the imaginary part $\boldsymbol{a}_{\Im,m}$.

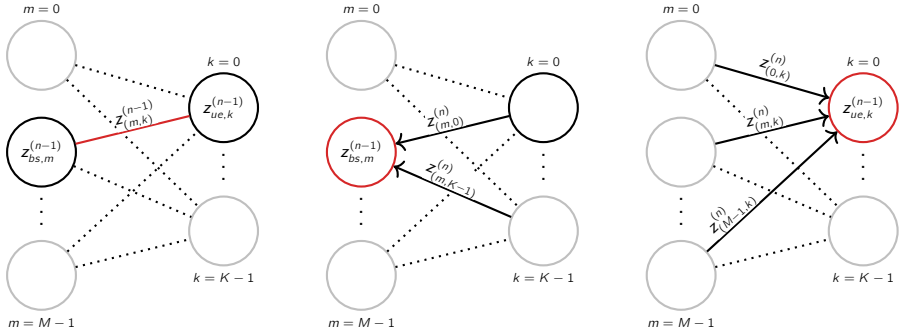
the labels. Given that these labels are expensive to obtain, unsupervised training is used. We interpret the output of the softmax layer as the probability each output level has to be the 'best' value. To select the best output labels we take the argmax of these probability vectors to select the maximum likelihood output levels, as depicted in (6.19).

The problem here is that the argmax function is non-differentiable. A common workaround is to use the argmax function in the forward pass of the NN and estimate the gradients using a softmax function in the backward pass. However, empirically, we found this approach to give poor results, as it often leads to greedy, sub-optimal solutions [113]. Motivated by the exploration-exploitation dilemma, we aim to introduce more exploration by treating the output of the NN as the parameters of a categorical distribution, from which samples are drawn rather than selecting the maximum value via argmax. This sampling introduces a degree of exploration, enabling the NN to observe a wider variety of (potentially negative) examples during training. This enables the model to escape sub-optimal local minima. By contrast, relying on the deterministic argmax function emphasizes exploitation, which risks converging to greedy, suboptimal solutions [113].

To train the NN in this way a sampling scheme of which the gradients can be estimated is required. We use a relaxed reparameterization gradient estimator known as the straight-through Gumbel-softmax estimator [113, 117]. To build up to this we discuss three variants: *(1)* the Gumbel-max estimator, *(2)* the Gumbel-softmax estimator and *(3)* the straight-through Gumbel-softmax estimator. First, we consider a reparameterization gradient estimator *(1)*. This estimator reparameterizes the sampling of a random variable by using a sample from a standard distribution combined with a deterministic transformation. In our case, this is referred to as the Gumbel-max trick, where a sample from the categorical distribution $l \sim \text{cat}(\boldsymbol{p}_{\Re,m})$ is obtained by sampling the standard Gumbel distribution and then transforming it. More concretely, sampling the categorial distribution is equivalent to adding samples from the standard Gumbel distribution to the logits and selecting the index with the highest value [113]

$$l = \underset{i\in\{0,\cdots,L-1\}}{\arg\max} \{[\boldsymbol{a}_{\Re,m}]_i + g_i\} \sim \text{cat}(\boldsymbol{p}_{\Re,m}) \tag{6.21}$$

where $g_i \sim \text{Gumbel}(0,1)$ are samples from the standard Gumbel distribution and $[\boldsymbol{a}_{\Re,m}]_i$ are unnormalized probabilities (i.e., the input to the softmax function as in (6.20)). However, for the Gumbel-max trick, the deterministic transformation is non-differentiable due to the argmax function. Therefore, a relaxed gradient estimator *(2)* is often used which replaces the argmax by a softmax function (typically with an added temperature variable τ). These soft samples are defined



(a) Edge feature update according to (6.23).

(b) Antenna node feature update following (6.26) and (6.25).

(c) User node feature update according to (6.27) and (6.24).

Figure 6.6: GNN overview: antenna nodes are on the left, user nodes on the right. Red indicates the element being updated, black the elements used to perform that update. The previous value of the element, before the update, is also used during the update.

as

$$S_{i,\tau} = \frac{e^{([\boldsymbol{a}_{\mathbb{R},m}]_i + g_i)/\tau}}{\sum_j e^{([\boldsymbol{a}_{\mathbb{R},m}]_j + g_j)/\tau}} \quad \text{for } i \in \{0, \cdots, L-1\}. \tag{6.22}$$

This approximates the (typically discrete) sample of the categorical distribution by a continuous random variable, which allows for the computation of gradients. This is known as the Gumbel-softmax trick. In the limit where $\tau \to 0$ this converges to a discrete (one-hot encoded) sample $\mathbb{1}_i^{|\mathcal{L}|}$. For some applications, this relaxation is acceptable, however, given that the aim of this work is quantization, a continuous relaxation would undo the desired quantization. To overcome this, we use the straight-through Gumbel-softmax estimator *(3)*. This estimator produces discrete samples in the forward pass and uses the gradients from the Gumbel-softmax trick in the backward pass. This can effectively be seen as using the Gumbel-max trick in the forward pass, and the Gumbel-softmax trick in the backward pass. An overview of this training procedure can be seen in Figure 6.5. Note that, the straight-through Gumbel-softmax estimator is only used during training. At inference time, the argmax function is used to select the output levels that are assigned the highest probability according to (6.19), to obtain the precoded quantized vector $\boldsymbol{y}$.

6.4.3 Neural Network Architecture

In this section, we discuss which type of NN architecture is adequate for non-linear precoding. For a formal discussion of how to select the NN architecture, based on statistical learning theory, we refer to Chapter 3 and [67, 64]. However, the main insight is as follows: by using certain NN architectures we restrict the NN to only learn certain groups of functions. As such, the hypothesis space of the NN is greatly reduced. This leads to more scalable NNs with fewer trainable parameters, which are easier to train and have lower computational complexity. The main idea lies in selecting the right NN architecture, which can learn a group of functions, which is known to cover the desired function. In this way, the reduction in expressivity does not affect the performance, as only functions outside of this group are ruled out. Next, we explain how this mechanism can be applied to the problem in this chapter by limiting the NN to only learn permutation equivariant/invariant functions.

For non-linear precoding, given the physical underlying system, the following equivariance and invariance properties need to be maintained. *i)* If the user order is permuted with permutation matrix $\boldsymbol{\Pi} \in \{0,1\}^{K\times K}$, the output of the precoding function should remain unchanged, hence $f(\boldsymbol{H}\boldsymbol{\Pi}, \boldsymbol{\Pi}\boldsymbol{s}) = \boldsymbol{y}_{\text{NL}}$. The learned function should thus be permutation invariant with respect to the symbol vector and columns of the channel matrix. *ii)* If the antenna ordering is permuted with permutation matrix $\boldsymbol{\Pi} \in \{0,1\}^{M\times M}$, the output of the precoding function should be permuted accordingly $f(\boldsymbol{\Pi}\boldsymbol{H}, \boldsymbol{s}) = \boldsymbol{\Pi}\boldsymbol{y}_{\text{NL}}$. The learned function should thus be permutation equivariant with respect to the rows of the channel matrix. When a NN imposes these properties on the learned functions, its hypothesis space is reduced as it can no longer learn all existing functions, but only the subset that respects these properties. Furthermore, this can only positively affect the learning performance, as the optimal function (i.e., a function that is permutation equivariant with respect to the antenna ordering, and permutation invariant with respect to the user ordering), is still covered by this restricted hypothesis space. This methodology is analogous to the linear precoding case in Chapter 3 where permutation equivariance with respect to the user and antenna dimensions was encoded in the GNN-based linear precoder.

Next, a GNN architecture is described which respects these properties. The proposed GNN is based on generalized message passing [68, 65]. The GNN learns the mapping from channel matrix and symbol vector to a precoded and quantized vector over a graph. The same graph as in the previous chapters is considered, i.e., it has a node for each transmit antenna at the BS, a node for each user and an edge between each antenna and user as can be seen in Figure 6.6a. The graph

is defined as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \mathcal{V}_M \cup \mathcal{V}_K$ is the set of nodes, with $\mathcal{V}_M$ the antenna nodes and $\mathcal{V}_K$ the user nodes, and $\mathcal{E}$ the set of edges, which contains all edges between the antenna and user nodes i.e., $(m, k) \in \mathcal{E}\ \forall m \in \mathcal{V}_M, k \in \mathcal{V}_K$. The graph is undirected, meaning that $(m, k) \in \mathcal{E} \leftrightarrow (k, m) \in \mathcal{E}$, as is the physical channel between each user and antenna. A general GNN performs a number of message-passing iterations. At iteration/layer n, a hidden representation is updated for each edge, antenna node, and user node, denoted by $\boldsymbol{z}_{(m,k)}^{(n)}$, $\boldsymbol{z}_{bs,m}^{(n)}$ and $\boldsymbol{z}_{ue,k}^{(n)} \in \mathbb{R}^{d_n}$ respectively. These representations are updated in layer n according to

$$\boldsymbol{z}_{(m,k)}^{(n)} = \sigma\left(\boldsymbol{W}_{\text{edge}}^{(n)} \boldsymbol{z}_{(m,k)}^{(n-1)} + \boldsymbol{W}_{\text{bs}}^{(n)} \boldsymbol{z}_{bs,m}^{(n-1)} + \boldsymbol{W}_{\text{ue}}^{(n)} \boldsymbol{z}_{ue,k}^{(n-1)}\right) \tag{6.23}$$

$$\boldsymbol{m}_{\mathcal{N}(k)} = \frac{1}{|\mathcal{N}(k)|} \sum_{m' \in \mathcal{N}(k)} \boldsymbol{z}_{(m',k)}^{(n)} \tag{6.24}$$

$$\boldsymbol{m}_{\mathcal{N}(m)} = \frac{1}{|\mathcal{N}(m)|} \sum_{k' \in \mathcal{N}(m)} \boldsymbol{z}_{(m,k')}^{(n)} \tag{6.25}$$

$$\boldsymbol{z}_{bs,m}^{(n)} = \sigma\left(\boldsymbol{W}_{\text{self,bs}}^{(n)} \boldsymbol{z}_{bs,m}^{(n-1)} + \boldsymbol{W}_{\text{neigh,bs}}^{(n)} \boldsymbol{m}_{\mathcal{N}(m)}\right) \tag{6.26}$$

$$\boldsymbol{z}_{ue,k}^{(n)} = \sigma\left(\boldsymbol{W}_{\text{self,ue}}^{(n)} \boldsymbol{z}_{ue,k}^{(n-1)} + \boldsymbol{W}_{\text{neigh,ue}}^{(n)} \boldsymbol{m}_{\mathcal{N}(k)}\right). \tag{6.27}$$

Here $\sigma(\cdot)$ is a non-linear activation and the matrices $\boldsymbol{W}_{\text{edge}}^{(n)}, \boldsymbol{W}_{\text{bs}}^{(n)}, \boldsymbol{W}_{\text{ue}}^{(n)}, \boldsymbol{W}_{\text{self,bs}}^{(n)}$, $\boldsymbol{W}_{\text{self,ue}}^{(n)} \in \mathbb{R}^{d_n \times d_{n-1}}$ and $\boldsymbol{W}_{\text{neigh,bs}}^{(n)}, \boldsymbol{W}_{\text{neigh,ue}}^{(n)} \in \mathbb{R}^{d_n \times d_n}$ are learned. The dimensions of these matrices (d_{n-1}, d_n) determine the number of features in layer $n-1$ and n respectively. Note that m denotes the antenna index, k the user index and (m, k) the edge between antenna m and user k. First, in (6.23) the edge features are updated according to the edge, antenna and user features from the previous layer. Second, in (6.24) message passing is performed from the neighbors of each user node to generate the user messages, in (6.25) message passing is performed from the neighbors of each antenna node to generate the antenna messages. Next, in (6.26) the antenna features are updated according to the antenna messages and the antenna features from the previous layer. Finally, in (6.27) the user features are updated according to the user features from the previous layer and the user messages. Figure 6.6 provides an overview of these computations. Note that these operations are repeated for a total of N layers: one input layer, one output layer and $N_h = N - 2$ hidden layers.

The inputs for edge (m, k), antenna node m and user node k in layer one of the GNN are defined in the following way

$$\mathbf{z}_{(m,k)}^{(0)} = \left[\Re\{[\mathbf{H}]_{m,k}\}, \Im\{[\mathbf{H}]_{m,k}\}\right]^{\mathsf{T}} \tag{6.28}$$

$$\mathbf{z}_{bs,m}^{(0)} = \left[0, 0\right]^{\mathsf{T}} \tag{6.29}$$

$$\mathbf{z}_{ue,k}^{(0)} = \left[\Re\{s_k\}, \Im\{s_k\}\right]^{\mathsf{T}}. \tag{6.30}$$

The outputs of the GNN are the precoded symbols, which are defined as the antenna node features in the final layer

$$\mathbf{z}_{bs,m}^{(N-1)} = \left[\Re\{y_m\}, \Im\{y_m\}\right]^{\mathsf{T}}. \tag{6.31}$$

Note, however, that the actual outputs of the GNN are probability vectors as discussed in the previous section $\mathbf{z}_{bs,m}^{(N-1)} = \left[\boldsymbol{p}_{\Re,m}, \boldsymbol{p}_{\Im,m}\right]^{\mathsf{T}}$. At inference time, these probability vectors are mapped to the precoded symbols according to (6.19).

6.5 Complexity Analysis

In this section, the complexity of the proposed GNN is computed in terms of real FLOPs.

6.5.1 Complexity of the GNN

The forward pass of the GNN is described in equations (6.23), (6.24), (6.25), (6.26) and (6.27). To compute the complexity, a distinction is made between the input layer, the four hidden layers ($N_h = 4$) and the output layer, as the dimensions of the inputs, the features, and the outputs are different in each of these layers. In the input layer $d_0 = 2$, in the hidden layers $d_n = d_h = 128\ \forall n \in \{1, \cdots, N-2\}$ and in the output layer $d_{N-1} = 2^{b+1}$. Additionally, (6.23) has to be computed for each edge (MK times), (6.24) and (6.27) for each user node (K times) and (6.25) and (6.26) for each antenna node (M times).

For the input layer, the number of real floating-point multiplications is

$$\begin{aligned}\mathcal{O}_{\text{in}}^{(\text{mul})} = &\underbrace{6MKd_h}_{(6.23)} + \underbrace{Kd_h}_{(6.24)} + \underbrace{Md_h}_{(6.25} \\ &+ \underbrace{2Md_h + Md_h^2}_{(6.26)} + \underbrace{2Kd_h + Kd_h^2}_{(6.27)}\end{aligned} \tag{6.32}$$

and the number of real floating-point additions is

$$\begin{aligned}\mathcal{O}_{\text{in}}^{(\text{add})} = &\underbrace{5MKd_h}_{(6.23)} + \underbrace{K(M-1)d_h}_{(6.24)} + \underbrace{M(K-1)d_h}_{(6.25)} \\ &+ \underbrace{Md_h + Md_h^2}_{(6.26)} + \underbrace{Kd_h + Kd_h^2}_{(6.27)},\end{aligned} \tag{6.33}$$

where the underbraces indicate the equations that require the indicated computations. One hidden layer requires

$$\mathcal{O}_{\text{h}}^{(\text{mul})} = \underbrace{3MKd_h^2}_{(6.23)} + \underbrace{Kd_h}_{(6.24)} + \underbrace{Md_h}_{(6.25)} + \underbrace{2Md_h^2}_{(6.26)} + \underbrace{2Kd_h^2}_{(6.27)} \tag{6.34}$$

real floating-point multiplications and

$$\begin{aligned}\mathcal{O}_{\text{h}}^{(\text{add})} = &\underbrace{3MKd_h^2 - MKd_h}_{(6.23)} + \underbrace{K(M-1)d_h}_{(6.24)} + \underbrace{M(K-1)d_h}_{(6.25)} \\ &+ \underbrace{2Md_h^2 - Md_h}_{(6.26)} + \underbrace{2Kd_h^2 - Kd_h}_{(6.27)}\end{aligned} \tag{6.35}$$

real floating-point additions. For the output layer, only the antenna node features are computed as they are the final output of the GNN. This requires

$$\mathcal{O}_{\text{out}}^{(\text{mul})} = \underbrace{6MKd_h2^b}_{(6.23)} + \underbrace{2M2^b}_{(6.25)} + \underbrace{4M2^{2b} + 2Md_h2^b}_{(6.26)} \tag{6.36}$$

real floating-point multiplications and

$$\begin{aligned}\mathcal{O}_{\text{out}}^{(\text{add})} = &\underbrace{6MKd_h2^b - 2MK2^b}_{(6.23)} \\ &+ \underbrace{2M(K-1)2^b}_{(6.25)} + \underbrace{4M2^{2b} + 2Md_h2^b - 2M2^b}_{(6.26)}\end{aligned} \tag{6.37}$$

real floating-point additions. In total, the GNN with $N_h = 4$ requires the following number of floating-point multiplications

$$\mathcal{O}_{\text{GNN}}^{(\text{mul})} = \mathcal{O}_{\text{in}}^{(\text{mul})} + N_h \mathcal{O}_{\text{h}}^{(\text{mul})} + \mathcal{O}_{\text{out}}^{(\text{mul})} \tag{6.38}$$

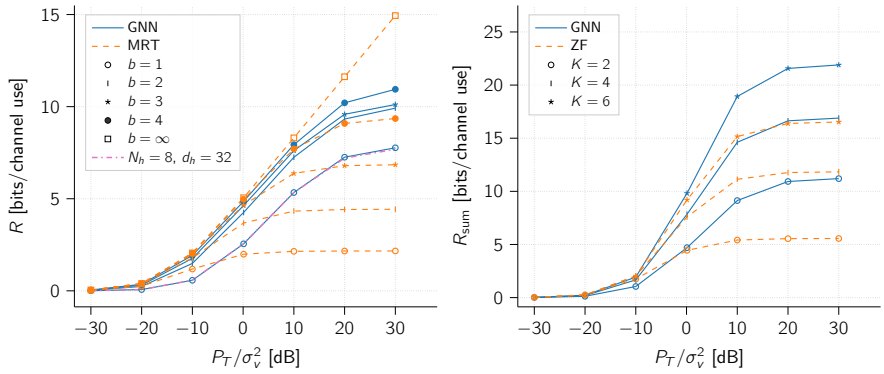
and the following number of floating-point additions

$$\mathcal{O}_{\text{GNN}}^{(\text{add})} = \mathcal{O}_{\text{in}}^{(\text{add})} + N_h \mathcal{O}_{\text{h}}^{(\text{add})} + \mathcal{O}_{\text{out}}^{(\text{add})} \tag{6.39}$$

This scales as $\mathcal{O}(d_h^2 MK + 2^b MK d_h + 2^{2b} M)$ multiplications and $\mathcal{O}(d_h^2 MK + 2^b MK d_h + 2^{2b} M)$ additions.

6.6 Simulation-based Performance Assessment

6.6.1 Training Parameters and Simulation Setup



(a) $M = 32,\ K = 1$ (b) $M = 32\ ,\ b = 1,\ K \in \{2, 4, 6\}$

Figure 6.7: Achievable rates averaged over the channel realizations taken from the test set. Comparing the GNN non-linear precoder with MRT or ZF, for the single-user case and a varying number of bits b (a) and the multi-user case $K \in \{2, 4, 6\}$ with one bit and a varying number of users (b).

For the following simulations, the total transmit power is normalized to $P_T = M$. During training, the SNR is fixed at $P_T/\sigma_v^2 = 20$ dB. All models are trained for 20 epochs unless specified otherwise. The batch size is set to 128, the Adam optimizer [61] is used with a learning rate of 5×10^{-3}. During training, the loss function in (6.13) is computed numerically. Hence, per channel, $N_s = 125$

complex Gaussian symbols are generated per user according to $s_k \sim \mathcal{CN}(0,1)$. These symbols are precoded and quantized for the current channel using the GNN. For this, non-uniform quantization levels are used, which are obtained using the Max-Lloyd algorithm, as described in Section 6.2.2. Next, the average power constraint is enforced by performing a scalar normalization over these $N_s = 125$ symbols according to (6.16). Finally, these precoded and quantized symbols ($\boldsymbol{y}_{\text{NL}}$), are used together with the intended transmit symbols ($\boldsymbol{s}$), to evaluate the expectation in the loss function according to (6.13). The training set consists of 200 000 generated channels, each of which is accompanied by $N_s = 125$ symbol vectors. The channels are sampled from a complex normal distribution $[\boldsymbol{H}]_{i,j} \sim \mathcal{CN}(0,1)$ if a Rayleigh fading channel is assumed. However, when constructing radiation patterns, a LoS channel is considered according to (6.10), where the user angle (in degrees) is randomly sampled from a discrete uniform distribution $\phi_k \sim \mathcal{U}\{0, 180\}$. All NN hyperparameters are selected on a validation set of 1000 channel realizations, while the simulation results are obtained on an independent test set of 10 000 channel realizations. The GNN used in this section contains one input layer, four hidden layers ($N_h = 4$), one output layer and $d_h = 128$ features are learned in the hidden layers unless specified otherwise. The non-linear activation function in all hidden layers is the leaky rectified linear unit (LReLU) which is defined as $\sigma(a) = \max(0, a) - 0.01\min(0, a)$. Finally, the temperature of the Gumbel-softmax estimator is set to $\tau = 1$.

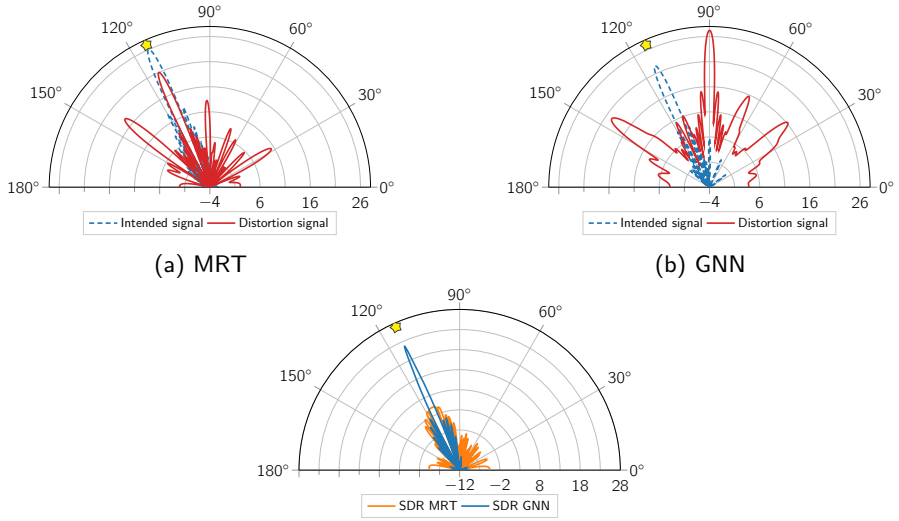
6.6.2 GNN-Based Non-Linear Precoding Single-User Case

In this section, the performance of the proposed GNN-based non-linear precoder is evaluated for the single-user case. In Figure 6.7a the rate is depicted for a varying SNR ($= P_T/\sigma_v^2$) for $M = 32$ transmit antennas, $K = 1$ user and a varying number of bits $b \in \{1, 2, 3, 4\}$. From this figure, it is clear that the GNN outperforms MRT precoding in the distortion-limited regime i.e., at high SNR. At high SNR, the GNN with one bit DACs delivers slightly better performance than MRT with three bit DACs. This highlights the ability of the GNN-based precoder to leverage over-the-air combining to provide a higher-quality signal. Next to this, the GNN outperforms classical linear precoding for one to four bit DACs. However, the performance gap decreases as the number of bits goes up. This is to be expected for two reasons: *i)* as the number of bits rises, the distortion goes down, hence there is less to gain by using quantization-aware precoding and *ii)* the problem complexity increases, as the number of possible output levels rises exponentially with the number of bits, this increased search space results in a harder problem to solve. In conclusion, in the single-user case, the proposed

method outperforms classical MRT precoding for one to four bits. However, the gains are most pronounced in the one-bit case.

In Figure 6.8, the radiation pattern of MRT is compared to the GNN for the one-bit single-user case $b = 1$, $K = 1$. This figure highlights that the GNN achieves its gain by sending the distortion in non-user directions, while the distortion in the user direction is kept to a minimum. This results in a higher SDR in the user direction as compared to MRT, as can be seen in Figure 6.8c. In Figure 6.9 the radiation pattern is depicted for the one-bit single-user case $b = 1$, $K = 1$ where the number of transmit antennas is varied $M \in \{8, 16, 32\}$. Given the increase in degrees of freedom when the number of transmit antennas grows, the GNN can control both the intended signal and distortion more precisely. This leads to an increase in SDR in the user direction as the number of transmit antennas increases.

In Figure 6.10a the transmitted symbols s are plotted in function of the estimated received symbols $\hat{s}$ for $M = 2$ transmit antennas, $K = 1$ user and $b = 1$ bit. In general, for complex channels and transmit symbols, after over-the-air combining, 2^{2bM} possible symbols can be received. These possible received signals are not uniformly distributed but depend on the channel realizations and the transmit signal at each antenna. Hence, selecting the right transmit signal at each antenna allows for control over the received signals. However, this is limited by the values of each channel realization. Nevertheless, Figure 6.10a illustrates that the GNN can more effectively utilize these 16 possible symbols after over-the-air combining to obtain a good estimate close to the transmitted symbol. This is illustrated by the difference in normalized mean squared error (NMSE), which is computed as NMSE $= \frac{\mathbb{E}\left(|s-\hat{s}|^2\right)}{\mathbb{E}(|s|^2)}$. The GNN achieves a NMSE of -10.11 dB, and MRT a NMSE of -7.43 dB. This increased performance scales with the number of transmit antennas as after over-the-air combining, 2^{2bM} possible symbols can be received. In Figure 6.10b the transmitted symbols s are plotted in function of the estimated received symbols $\hat{s}$ for $M = 32$ transmit antennas, $K = 1$ user and $b = 1$ bit DACs. For MRT, the estimated symbols differ largely from the transmitted symbols. When using GNN-based non-linear precoding a remarkable improvement can be observed. This is expected given the increase in transmit antennas, which leads to $2^{2bM} = 2^{64}$ possible received symbols. This demonstrates the potential of using many transmit antennas in combination with one bit DACs. However, the challenge lies in selecting the right combination of DAC output levels at each transmit antenna to obtain the desired received signal. Figure 6.10b demonstrates the ability of the GNN to select the correct output levels at each DAC which results in estimated symbols that match very closely the transmit symbols. To quantify this, the NMSE is computed over all noiseless channel realizations and symbols



(a) MRT

(b) GNN

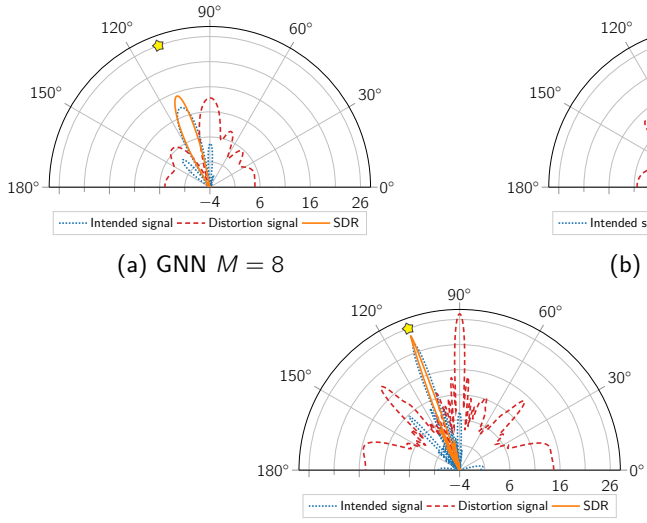
(c) SDR

Figure 6.8: Radiation pattern of the intended signal $P_{\text{lin}}(\tilde{\phi})$ and distortion signal $P_{\text{dist}}(\tilde{\phi})$ [dB] for a 1-bit DAC, $M = 32$ and $K = 1$. The GNN precoder (a) and MRT precoding (b) are compared for a pure LoS channel and a half-wavelength ULA. The user is indicated by a star. In (c) the signal-to-distortion ratio (SDR) radiation pattern $P_{\text{SDR}}(\tilde{\phi})$ is depicted [dB].

Table 6.1: NMSE for MRT and GNN on all (noiseless) channels and symbols from the test set, for $M = 32$, $K = 1$.

Precoder	$b = 1$	$b = 2$	$b = 3$	$b = 4$
NMSE MRT [dB]	-5.38	-13.02	-20.31	-27.70
NMSE GNN [dB]	-22.99	-29.43	-29.87	-32.52

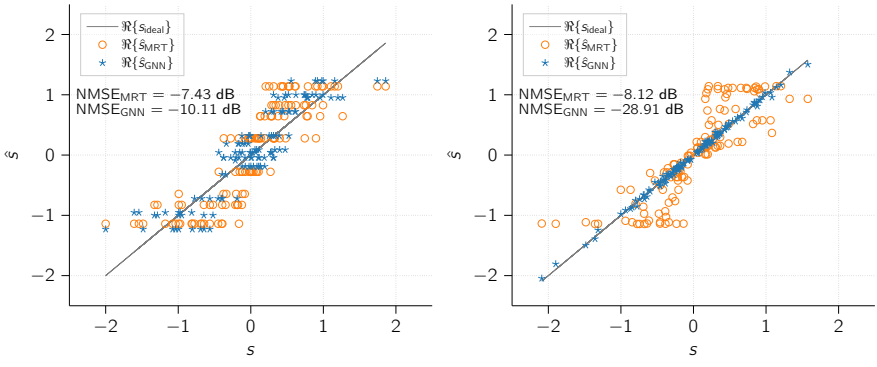
from the test set. This is done for MRT and GNN precoding for $b \in \{1, 2, 3, 4\}$ bits in Table 6.1. For one-bit DACs the GNN achieves a NMSE of -22.99 dB, while MRT leads to a NMSE of -5.38 dB, highlighting the performance gap between MRT and the GNN. As the number of bits increases, this performance gap decreases. However, even for $b = 4$ bits, the GNN achieves a NMSE of -32.52 dB, while MRT precoding achieves a NMSE of -27.70 dB.



(a) GNN $M = 8$ (b) GNN $M = 16$

(c) GNN $M = 32$

Figure 6.9: Radiation pattern of the GNN for the intended signal $P_{\text{lin}}(\tilde{\phi})$ and distortion signal $P_{\text{dist}}(\tilde{\phi})$ [dB] for a 1-bit DAC, varying number of transmit antennas $M \in \{8, 16, 32\}$ and $K = 1$ user. The user angle is $\phi = 110°$.



(a) $M = 2$ (b) $M = 32$

Figure 6.10: Real part of the estimated symbol $\hat{s}$ in function of the transited symbol s, over a single realization of a noiseless Rayleigh fading channel. Comparing MRT with GNN precoding, for $M = 2$ (a) and $M = 32$ (b), $K = 1$ and $b = 1$.

6.6.3 GNN-Based Non-Linear Precoding Multi-User Case

In this section, the proposed precoder is evaluated in the multi-user scenario. In Figure 6.7b the sum rate in function of the SNR is depicted for $M = 32$ transmit antennas, $b = 1$ bit and a varying number of users $K \in \{2, 4, 6\}$. This figure shows that the GNN outperforms ZF in the distortion-limited regime, i.e., at high SNR. As the number of users increases, the distortion receives less of a beamforming gain, and as such becomes more spatially spread out. Consequently, there is less to gain in terms of system performance. Additionally, this reduced gain is explained by the fact that the more users are present, the harder it is to train the GNN.

In Figure 6.11, the sum rate is illustrated in function of the number of bits for a fixed SNR$= P_T/\sigma_v^2 = 20$ dB, for $M = 32$ transmit antennas and $K \in \{1, 2, 4, 6\}$ users. When comparing the GNN precoder with the baseline (MRT or ZF) it is clear that the proposed method is most performant when few users are present and few bits are considered. When more users are present and/or more bits are considered, the proposed method shows similar or lower performance than the baseline. This is expected as the potential gains diminish as the number of users increases, and as the complexity of the problem exponentially increases with the number of bits. Note that in Figure 6.11 two GNNs are considered, the standard model with $d_h = 128$ features and a bigger model with $d_h = 256$ features. When the problem complexity is higher, i.e., more bits and users are considered, it is shown that the larger GNN ($d_h = 256$) achieves better performance as compared to the smaller model ($d_h = 128$) and ZF precoding. However, this comes at the cost of an increased model size and complexity. Nevertheless, previously the most critical case was identified as the one with few users and few bits, in this regime, the proposed smaller model ($d_h = 128$) outperforms the baseline.

6.7 Power Consumption

To quantify the power consumption of the proposed precoder, the consumed power of the DACs and the processing due to the GNN are taken into account.

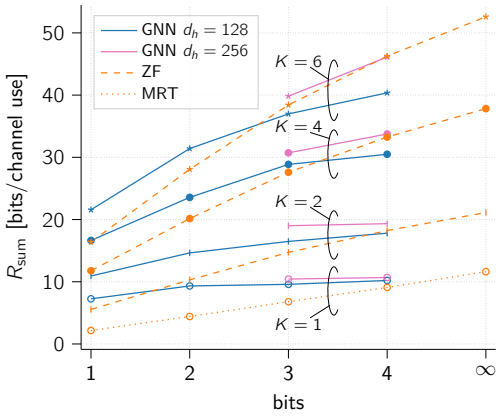


Figure 6.11: Achievable rates on test set, $P_T/\sigma_v^2 = 20$ dB. Comparing GNN and MRT or ZF for $M = 32$, $K \in \{1, 2, 4, 6\}$. A bigger GNN ($d_h = 256$) is added for $b \in \{3, 4\}$.

6.7.1 DACs Power Consumption

We consider the widely used current steering DAC architecture, with the following power consumption model [13]

$$P_{\text{DAC}} \approx \frac{1}{2} V_{dd} I_0 (2^b - 1) + b C_p \frac{f_s}{2} V_{dd}^2 \tag{6.40}$$

where V_{dd} is the supply voltage, I_0 the current of the current source of the least significant bit and C_p the parasitic capacitance of each transistor. Based on [13] we use $V_{dd} = 3\,\text{V}$, $I_0 = 10\,\mu\text{A}$ and $C_p = 1\,\text{pF}$.

For a fair comparison of the consumed power, a fixed quality of service is selected at which the proposed precoder is compared to the conventional precoder. As a concrete example, a rate constraint of 7 bits/channel use is considered for the scenario in Figure 6.7a at an SNR = 20 dB. Assuming a roll-off factor of $\alpha_{rol} = 0.1$, this corresponds to a constraint on the spectral efficiency of $\eta_{\text{SE}} = \frac{R}{1+\alpha_{rol}} \geq 6.4$ bit/s/Hz. Note that this achievable rate and spectral efficiency assume a normalized large-scale fading gain of one and do not account for path loss or shadowing effects. From Figure 6.7a, it is clear that the GNN can meet this constraint by using DACs with $b = 1$ bits while MRT requires $b = 3$ bits. Given the required bit resolution and sampling frequency, the consumed power of all DACs is $P_{\text{DACs,tot}} = 2MP_{\text{DAC}}$. Note that we here consider two types of DACs, namely the classical baseband DAC, whereby the system relies on analog upconversion to place the signal at the desired carrier frequency, and the RF-DAC

which directly synthesizes the signal at the desired carrier frequency. For the classical baseband DAC we consider a sampling frequency of $f_s = 4B$, where B is the bandwidth and an oversampling factor of four is applied. For the RF-DAC we select $f_s = 4f_c/(2n-1)$, which places the baseband signal at the center of the n^{th} Nyquist zone. In practice, the second Nyquist zone ($n = 2$) is most commonly utilized [87]. Note that for the RF-DAC the sampling frequency is independent of the bandwidth, however, the condition $f_s \geq B$ must be satisfied, which is generally the case, hence $f_s = \frac{4}{3}f_c$. The consumption of the baseband DACs is illustrated in Figure 6.12 in function of the bandwidth B. The proposed precoder allows the DACs to consume a factor of 4-7 less power. In Figure 6.13 the power consumption of the RF-DACs is given as a function of the bandwidth. Note that this consumption is constant over the bandwidth as the sampling frequency for the RF-DAC is linked to the carrier frequency $f_c = 3.5$ GHz, rather than the bandwidth. The proposed precoder allows the RF-DACs to consume a factor of 3 less power as compared to MRT. This is to be expected as the power consumption of the RF-DACs is dominated by the dynamic consumption (term 2 in (6.40)), which scales linearly with the number of bits, while the sampling frequency is fixed for both precoders.

6.7.2 GNN Processing Power Consumption

As computed in Section 6.5, the total number of FLOPs required to run the GNN one time is $\mathcal{O}_{\text{GNN}}^{(\text{add})} + \mathcal{O}_{\text{GNN}}^{(\text{mul})}$. However, given that non-linear precoding is considered, the GNN needs to be executed for each transmit symbol. Hence, to compute how many forward passes are required per second, the symbol rate is computed. Consider a transmission bandwidth $B = (1 + \alpha_{\text{rol}})/T$, where T is the symbol period and $\alpha_{\text{rol}} = 0.1$ a roll-off factor. This gives a symbol rate of $R_s = \frac{B}{1+\alpha_{\text{rol}}}$ symbols per second. Given this, the GNN needs to precode R_s symbols per second. Hence the total time complexity of the GNN is $R_s(\mathcal{O}_{\text{GNN}}^{(\text{add})} + \mathcal{O}_{\text{GNN}}^{(\text{mul})})$ FLOPs per second. Given this, we can compute the required speed of the accelerator given a certain bandwidth. For instance, when considering $B = 1$ MHz the considered accelerator requires a speed of 12 TFLOPs/second. Based on the NN accelerator comparison from [118], we select the accelerator from [119] that achieves the speed requirement of 12 TFLOPs/s and has a high efficiency of $\eta = 646.6$ TFLOPs/s/W. This accelerator operates on 8-bit floating point numbers while the current GNN is trained on 32-bit floating point numbers. The influence of this reduced precision on the performance of the GNN should be further investigated. Nevertheless, to get an estimate of the achievable power consumption of the GNN processing, it is expressed in function of the bandwidth

as $P_{\text{GNN}} = \frac{1}{\eta}\frac{B}{1+\alpha_{\text{rol}}}\left(\mathcal{O}_{\text{GNN}}^{(\text{add})} + \mathcal{O}_{\text{GNN}}^{(\text{mul})}\right)$. First we consider the baseband DACs. From Figure 6.12 it can be noted that the proposed solution provides a power reduction when the system bandwidth is below 200 kHz. When the bandwidth is further increased, the gain of using fewer bits in the DACs is overshadowed by the increased GNN processing power consumption. However, by noting that the computational complexity of the GNN in (6.38) and (6.39) scales linearly with the number of layers N_h but quadratically with the number of features d_h, the processing power consumption can be reduced. By changing the depth and width of the GNN to $N_h = 8$, $d_h = 32$ rather than $N_h = 4$, $d_h = 128$, the performance (for $K = 1$, $b = 1$) is maintained while the power consumption is drastically reduced, as can be seen in Figure 6.7a and Figure 6.12. This results in a power reduction for a system bandwidth up to 3.5 MHz. Note that the speed requirement for the GNN with $N_h = 8$, $d_h = 32$ is still met, as the selected accelerator from [119] has a maximum speed of 39.8 TFLOPs/s, while at $B = 4$ MHz the GNN, requires 8.82 TFLOPs/s. Next to this, we study the power consumption when considering RF-DACs in Figure 6.13. The consumption of RF-DACs is much larger as compared to baseband DACs, hence the added consumption of the GNN processing is minor as compared to the RF-DAC consumption. This leads to power savings, which are maintained for much higher bandwidths. However, in this situation, the limiting factor is the speed of the NN accelerator rather than its power consumption. The accelerator from [119] has a maximum speed of 39.8 TFLOPs/s. This limits the maximum bandwidth when using the GNN with $N_h = 4$, $d_h = 128$ to 1.95 MHz[3] while the maximum achievable bandwidth for the GNN with $N_h = 8$, $d_h = 32$ is 15.8 MHz. Nevertheless, over the 15.8 MHz bandwidth, a power reduction of 2.9 is maintained. As a perspective, future work could investigate more in-depth the performance and power consumption trade-off for the optimal GNN structure.

Additionally, it is worth noting that, the current analysis only considers the increase in GNN power consumption and the decrease in DACs power consumption, while other factors might be at play. For instance, the reduced resolution could have auxiliary effects as the power consumption in the digital chain can be reduced when this lower resolution is exploited, e.g., the power consumption of digital filters can be reduced as fewer bits need to be computed. Moreover, the power consumption of the digital fronthaul, which connects the baseband unit with the remote radio units, can also be reduced when fewer bits are required. This could allow a large energy reduction, especially in a distributed/cell-free scenario where long fronthaul connections may connect central processing units to access points. These indirect effects should be considered in a further analysis.

[3]Note that for completeness the expected power consumption for higher bandwidths is also added to Figure 6.13.

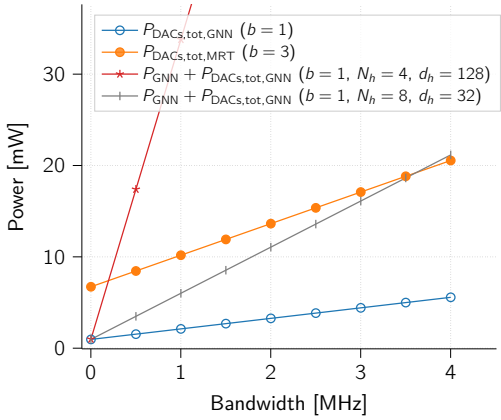


Figure 6.12: Power consumption of the DACs and GNN processing in function of the bandwidth.

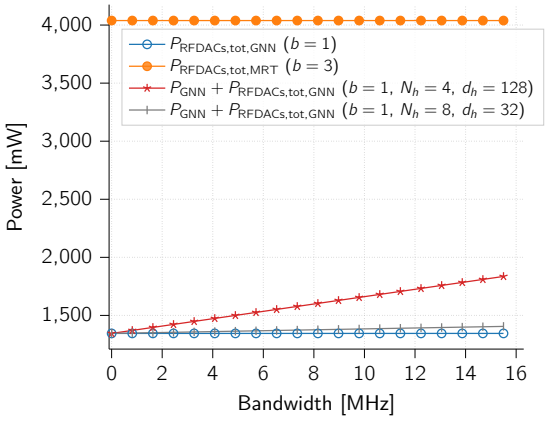


Figure 6.13: Power consumption of the RF-DACs and GNN processing in function of the bandwidth.

6.8 Practical Implementation Challenges

While the proposed methods are promising in terms of system-level performance, practical deployment raises several challenges. Given that the proposed method relies on non-linear precoding, the computational complexity is linked to the symbol rate and hence, the system bandwidth. Consequently, the current system is limited to bandwidths of 15.8 MHz due to the limited processing speed of the current state-of-the-art hardware accelerator. However, given the rapid advancements in hardware accelerators driven by growing demand for NN applications, the speed and efficiency of these hardware accelerators are expected to improve further [118]. Furthermore, training and deployment raise several challenges. The base model could be trained on simulation data (generated channels in our case). However, the distribution of the real-life measured channels will be different from the simulated channels. This phenomenon is known as covariate shift, as was outlined in Chapter 4 for the PA precoding case. As discussed in Chapter 4, by training in a simulated environment, a good initialization point of the model can be obtained, which can then be further fine-tuned using a limited set of measured real-life data [82]. By pre-training the model on simulation data, the amount of real-life measurements, which are often expensive and time-consuming to collect, can be limited [82]. Once deployed, another challenge arises, namely, maintaining the model when the data distribution varies over time. For instance, the properties of the wireless environment can change due to changes in the physical environment. Hence, the model needs to be continually updated. However, a trade-off arises here to avoid the problem known as catastrophic forgetting, where updating the model on new data usually results in a severe performance degradation on the old data [120]. To overcome these challenges, we refer to the rich literature on continual learning [120]. Finally, the proposed models should be compliant with current standards. As it stands, machine learning models are not standardized [121]. However, standardization instances have reported that testability and consistent system behavior are concerns [121]. Hence, these models should be extensively tested to meet current standards. More specifically, related to this work practical constraints such as the ACLR might be problematic due to the non-linear behavior of the DACs. Such constraints could be incorporated into the loss function, during training, to be compliant with the current standard.

6.9 Conclusion

In this chapter, a coarsely quantized downlink massive MIMO system is considered. A GNN is proposed, which leverages the many antennas at the BS to transmit the

non-linear quantization distortion in non-user directions. By doing so, a significant increase in achievable sum rate is obtained. For instance, when considering one-bit quantization in the single user case at high SNR, a three times increase in achievable rate as compared to classical precoding is obtained. This allows for using DACs with fewer bits to achieve a similar sum rate as traditional precoding techniques, which require more bits. The power consumption of baseband DACs is reduced by a factor of 4-7. However, when considering the additional processing power consumption, the total power reduction is only valid for bandwidths up to 3.5 MHz. Additionally, the power consumption of RF-DACs is reduced by a factor of 3. However, due to their higher absolute power consumption, the impact of the additional processing power consumption is limited, as the total power consumption is still reduced by a factor of 2.9. Next to this, a more holistic power consumption analysis, including the full digital chain and fronthaul consumption, can further improve these results. Moreover, future work should address the high processing power. For this two options remain open. *i)* The proposed solution can be further optimized by studying smaller GNN architectures, combined with model pruning and quantization of the model weights and activations. By doing so, the computational complexity of the current model could be significantly reduced. *ii)* The proposed solution relies on non-linear precoding. As such, the computational complexity scales linearly with the system bandwidth, which is the critical issue concerning the high processing power. Although challenging, future work could investigate this issue by using linear precoding and/or reduced-complexity non-linear precoders, which significantly reduce the computational complexity. Finally, the impact of channel estimation errors on the performance of the proposed method should be investigated.

Chapter 7

Valorization

In this chapter, the potential valorization of the research is discussed. The focus lies on how learning-based precoding for massive multiple input multiple output (MIMO) under hardware nonlinearities can contribute to more energy-efficient wireless networks and how this aligns with economic and societal objectives. Finally, the opportunities and limitations of artificial intelligence (AI)/machine learning (ML) for physical layer processing are discussed. We discuss the barriers and enablers towards deployment of the AI/ML-based solutions proposed in this thesis. This is done for four different facets, namely: societal, standardization, regulation and technical.

7.1 Economic and Ecological Alignment

The ecological motivation for this work is grounded in the growing carbon footprint of the information and communication technologies (ICT) sector and the explicit sustainability ambitions expressed in, for instance, the European Green Deal and the Sustainable Development Goals (SDGs). The wireless communications sector is currently falling behind these ambitions. If traffic growth continues at current rates, even maintaining the present level of energy consumption requires drastic improvements in energy efficiency. The work in this thesis contributes to these ambitions by focusing on techniques that reduce power consumption in the most energy-intensive parts of the base station.

Next to this, it is evident that future-proof wireless networks are essential for

our economy in the coming decades. This has been explicitly formulated by the European Commission in its strategy for data [122] and by the government of Flanders, that stipulated 'Digital society 2025' as a prioritized transition area in its reform program [123]. Energy efficiency of wireless networks, in terms of consumption for a given amount of data, has been improved over subsequent generations. However, in absolute terms, as a consequence of the mobile data traffic growth, the energy consumption has gone up. We should recognize that economic factors and full coverage guarantees have prevailed over ecological concerns in the past. Operators traditionally did not consider the power consumption of the base stations (BSs) as a major cost factor. Moreover, the maximum output power of equipment has been considered a quality factor, the higher the better[1]. Hence, there was little incentive for equipment vendors to focus innovation on the reduction of maximum power amplifier (PA) output power or digital-to-analog converter (DAC) consumption. However, now, the time has come when economic and ecological goals can and will reinforce each other. There are several reasons for this. First, energy consumption has become an important factor in the operating expenses of wireless networks, which is further expected to worsen. Indeed, both the growth in data and the increased price of energy contribute to this. Consequently, for operators, energy reduction directly results in cost reduction. Second, the cost of the PAs is an important component in the bill of materials (BOM) of wireless access equipment [7]. This cost is highly related to the output power. Thirdly, reducing carbon footprint has become a commercial value in itself. It could become an important criterion for customers, as is the 'green energy' argument for energy suppliers. Moreover, penalties and costs associated to CO2 emissions will apply.

From an economic perspective, both CAPEX and OPEX are impacted. Operating PAs closer to saturation allows for lower maximum output powers, smaller heat sinks and less demanding cooling systems, which can reduce the BOM for radio equipment. Similarly, reducing the DACs resolution lowers their power consumption and potentially their cost, while also decreasing the required fronthaul capacity and associated equipment.

The strategic dimension of this work can be related to concrete application domains and stakeholders. We identify the potential strategic importance of the pursued innovation for different players in the value chain and in different sectors:

1. Equipment vendors, who design and manufacture BSs. The main players here are Ericsson, Huawei, and Nokia. For instance, Ericsson is pursuing

[1]This can be compared to the horsepower of a car being perceived as a measure for how fast it can go, while many other factors come to play, and speed limits apply.

energy efficiency with a high priority as stated by their CTO: "From both cost and carbon footprint perspectives, energy is one of the industry's biggest challenges" [124]. The proposed precoding techniques enable more aggressive use of low-cost, low-power RF components, especially if the output power requirement of individual PAs can be reduced to levels at which CMOS implementations become feasible. In the long term, this could open the possibility to build base stations using similar low-power components as in handsets. The graph neural network (GNN)-based precoders presented in this thesis can support such architectures by managing the nonlinearity and distortion of these less ideal PAs and low-resolution DACs.

2. Mobile/wireless network operators, either public or private. The presented energy-saving techniques result in reduced OPEX and potentially lower cooling and infrastructure costs, without sacrificing quality of service.

3. Companies and entities installing and deploying a high-performance wireless (private) network as a crucial asset for their operation. This includes companies in diverse sectors beyond the traditional ICT sector, such as manufacturing sites (supporting Industry 4.0) and logistic hubs (e.g., ports), and health or educational campuses. For instance, Brussels airport [125] and the port of Antwerp [126] were the first to implement 5G in Belgium as a stand-alone network. Solutions that can provide the same quality of service (QoS) with less consumption will incur reductions both in CAPEX and in OPEX.

Rising energy prices amplify the relevance of these application domains. In regions with high electricity costs or where carbon pricing is implemented, the monetary value of reducing PA and DAC power consumption becomes substantial over the lifetime of the equipment.

7.2 AI Potential and Limitations in the Physical Layer

Recently, the application of AI/ML for wireless communications has garnered considerable attention [127]. In relation to this thesis, the application of AI/ML to physical layer communications has also experienced a surge in research interest [128, 129, 130]. This increased interest has led to high expectations regarding the potential impact on the physical layer. The work in this thesis provides a nuanced view of what is possible by showcasing both successful applications and clear

limitations. Next, we discuss the barriers and enablers towards deployment of the AI/ML-based solutions proposed in this thesis. This is done for four different facets, namely: societal, standardization, regulation and technical.

Societal Societal factors such as ecology, autonomy and privacy can act both as enablers and barriers toward deployment of AI at the physical layer. The main enabling factor is the environmental gain associated with the energy savings as envisioned in this thesis, as discussed in the previous section. Next to this, the methods developed in this thesis allow the exploitation of inexpensive hardware by mitigating associated distortion effects. This enables more affordable infrastructure which can accelerate deployment. However, the use of AI at the physical layer also is associated with a number of societal barriers. For instance, these models require extensive data collection and model updates, which raises privacy concerns. This is further discussed in the regulations paragraph. Another factor at stake is fairness. Given the continual model updates the behavior of these learned precoders could be hard to guarantee/predict. If not properly managed this could introduce fairness issues. For instance, if optimization is based on a single metric, predictable coverage and QoS in all environments could be hard to guarantee. This may reduce trust from operators and regulators.

Standardization In cellular systems, standardization is a crucial mechanism to ensure global interoperability, predictable performance, and regulatory compliance across vendors and operators. Looking to 6G, the integration of native AI/ML into the radio access network (RAN) and even the physical layer is envisioned, which creates both an opportunity and a requirement to frame neural network based precoding within future standardization efforts [131, 132]. Given the large push to integrate AI into wireless networks by major industry players and international bodies such as the international telecommunication union (ITU), initial standardization activities around AI-native and AI-assisted networks are emerging. However, these efforts are still largely fragmented across different standardization organizations and work items [131]. For instance, different frameworks for a standardized integration of AI/ML into wireless networks have been developed by third generation partnership project (3GPP), ITU-T, O-RAN and European telecommunications standards institute (ETSI). On the one hand, this push by different organizations enables the integration of AI/ML into the current standard. On the other hand, standardizing AI/ML is often quite challenging and can form a barrier. For instance, the developed frameworks by 3GPP, ITU-T, O-RAN and ETSI predominantly take a management- and lifecycle-oriented view: they define how ML models are trained, validated, deployed and monitored in

operational networks, rather than prescribing specific model architectures or use cases at the physical layer [131]. A common focus across these frameworks is on performance monitoring, trustworthiness, and validation of AI/ML models as testability and consistent system behavior are key concerns. For example this is done through explicit lifecycle phases (training, emulation, deployment, inference), model metadata and trust indicators, and continuous key performance indicator (KPI)-based monitoring and retraining triggers [131]. While this provides essential structures for integrating neural-network precoders into future systems, it also implies that the burden of defining physical layer specific requirements, test cases, and performance guarantees for AI-based precoding remains largely with individual working groups and implementers, contributing to the currently fragmented landscape [131].

If we look at the initial development of the 6G standard we see that it has been agreed upon that orthogonal frequency division multiplexing (OFDM) will be the major waveform [133]. This waveform is compatible with the developed precoder from Chapter 5. Next to this, we look at the considered frequency bands for both 5G and 6G. In 5G two frequency bands are considered, namely frequency range (FR)1 (410 - 7125 MHz) and FR2 (24.25 - 52.6 GHz). Additionally, we see that FR3 (7.125 - 24.25 GHz) is being viewed as the most likely frequency band for 6G [134, 135]. In this thesis, operation in FR1 was assumed as in line with the 5G standard. Increasing the frequency from the traditional sub-6GHz band to either FR2 in 5G or potentially FR3 in 6G, increases non-linear behavior in PAs [136]. Additionally, due to the increased path loss at higher frequencies, driving PAs with a larger gain is beneficial. Both these factors can be addressed by the distortion-aware precoders as developed in Chapter 3 and Chapter 5. However, moving up in carrier frequency also brings about associated challenges. Increasing the carrier frequency results in a reduced coherence time. This creates more stringent constraints on the inference time that is available for precoding. Hence, the complexity and inference time of the developed precoders might become barriers for deployment. Directly linked to this are the considered bandwidths for both 5G and 6G. In 5G bandwidths of 5-100 MHz are currently considered for FR1 and 50-400 MHz for FR2. For future 6G deployments bandwidths have not been agreed upon. However, considering the potential move to FR3, this will result in higher bandwidths as compared to FR1. Even the considered bandwidth at FR1 ($\leq$ 100 MHz) is challenging for the nonlinear precoders proposed in Chapter 6 as it was illustrated that inference is only feasible for bandwidths up to 15.8 MHz. When considering FR2 in 5G or FR3 in 6G this becomes even more problematic as the bandwidth further increases. Consequently, non-linear symbol level precoding as outlined in Chapter 6 hits a clear barrier in terms of inference complexity for these bandwidths.

Regulation Beyond standardization, regulatory aspects play an important role in determining whether AI/ML-based precoding can be deployed in operational wireless networks. The use of ML algorithms at the physical layer is currently not addressed by any telecom-specific regulations targeting AI as such. Instead, regulatory constraints arise indirectly through existing regulations concerning radio emissions, equipment safety, data protection, and the use of AI in safety-critical systems.

A first and immediate regulatory constraint relevant to this thesis concerns radio emission limits, in particular the adjacent channel leakage ratio (ACLR) and related out-of-band emission requirements defined by regulators and enforced through standards such as those of 3GPP and ETSI [137]. Since the proposed precoding schemes intentionally operate PAs and DACs in nonlinear regimes, they will increase spectral regrowth if not properly controlled. From a regulatory perspective, this is not fundamentally incompatible with deployment, provided that compliance with emission masks can be guaranteed. This can be addressed either by explicitly incorporating ACLR-related constraints into the training objective, or by dynamically adapting the precoder behavior depending on spectrum usage. In scenarios where adjacent bands are unused or when the system operates at low load using only central subcarriers, temporary relaxation of out-of-band emission constraints may be acceptable, whereas stricter precoding strategies must be enforced when adjacent-band receivers are active or when the load is high [11].

More broadly, regulations regarding the use of AI in Europe are mandated through the European Union Artificial Intelligence Act [138]. This framework adopts a risk-based approach and does not prohibit the use of AI for wireless communications or physical layer signal processing. Learning-based precoders would generally fall into the category of low- or limited-risk AI systems, as they do not directly interact with users, perform biometric identification, or make decisions affecting fundamental rights [139, 140]. As a result, no prior authorization or conformity assessment specific to AI functionality is expected solely due to the use of neural networks in the precoder. However, obligations related to transparency, documentation, robustness, and risk management may apply at the system level, especially when AI components influence network performance or reliability.

In addition, existing regulatory frameworks for radio equipment, such as the electromagnetic compatibility (EMC) directive in Europe, require manufacturers to demonstrate that equipment operates safely, efficiently uses the radio spectrum, and is robust against interference [141]. From this perspective, AI-based precoding is treated as an implementation choice within the transmitter, rather than as a distinct regulatory category. This implies that neural network-based precoders must ultimately be validated through conventional compliance testing. The non-

deterministic nature of learned models can therefore be perceived as a regulatory barrier, not because it is prohibited, but because it complicates verification, testing, and certification procedures that were historically designed for deterministic signal processing chains.

Finally, while neural network (NN)-based precoding inference does not typically involve personal data, training and retraining mechanisms are data-driven. If channel measurements, network telemetry, or user-related information are stored or processed for (re)training or fine-tuning, data protection regulations such as the general data protection regulation (GDPR) become relevant [142]. This primarily affects how training data are collected, anonymized, and retained, rather than the inference-time operation of the precoder. As such, regulatory compliance is largely a system-level and operational concern, rather than a fundamental obstacle to the techniques proposed in this thesis. Nevertheless, this could introduce extra friction which might slow down adoption.

Overall, regulation does not prevent the deployment of AI-based precoding at the physical layer, but it imposes constraints on emissions, robustness, testability, and operational transparency. The non-deterministic behavior of these precoders increases the burden of validation, compliance testing and certification, which may slow down regulatory acceptance.

Technical From a technical perspective, the deployment of AI/ML-based precoding at the physical layer faces a number of both enabling factors and barriers. One key challenge concerns the mismatch between training and deployment conditions. As training is typically performed offline under assumed channel and hardware models, the distribution encountered during online operation may differ due to model mismatch, changes in propagation conditions, mobility patterns, or hardware aging. This phenomenon, known as covariate shift, was investigated in Chapter 4. While fine-tuning on representative data can mitigate such effects, long-term operation in dynamic environments may require periodic retraining or adaptive updates. As a result, mechanisms for continual learning, as well as robust performance monitoring and fallback strategies, are important for reliable deployment.

A second technical barrier is related to scalability. Modern and future wireless systems continue to increase the number of transmit antennas, users, and subcarriers, which directly impacts the dimensionality of the precoding problem. Although the GNN-based architectures proposed in this thesis exploit permutation equivariance to reduce the hypothesis space, the computational complexity still grows with system size. In particular, it should be studied how the model size of

GNN-based precoders scales with respect to the number of transmit antennas and subcarriers. Given that increasing the number of transmit antennas and subcarriers typically results in a more complex optimization problem, we expect that the model size will not scale linearly with the number of antennas and/or subcarriers. Identifying these scaling laws and designing appropriate schemes to manage the computational complexity remain important topics for future research.

Closely related to scalability is the trade-off between performance gains and computational complexity. As discussed in previous chapters, the energy savings achieved through distortion-aware precoding must be balanced against the additional power consumption of GNN inference. Linear GNN-based precoders benefit from operating at the channel coherence time scale, resulting in a manageable increase in inference complexity and energy cost. In contrast, non-linear, symbol-level precoding is more demanding. This limits its practical attractiveness to specific operating regimes. In particular, such approaches are most viable in scenarios with severe hardware constraints or limited bandwidth, where the relative performance gains outweigh the inference cost.

Another technical consideration concerns integration into existing signal processing chains and operational workflows. Deploying learning-based precoders in practice requires not only efficient inference, but also mature machine learning operations (MLOps) pipelines that support model training, validation, deployment, monitoring, and rollback [131]. Unlike traditional signal processing algorithms, neural network-based components introduce additional system complexity in terms of software dependencies, version control, and performance assurance. Ensuring stable and predictable behavior across software and hardware updates is therefore a non-trivial technical requirement, which is in line with current standardization efforts for AI in telecommunications [131].

Overall, the technical analysis in this thesis indicates that NN-based precoding is most promising when applied in a targeted and well-defined scope. Linear, distortion-aware GNN-based precoders that operate at the channel level offer a favorable balance between performance, complexity, and robustness, making them suitable for near-term deployment. More aggressive non-linear, symbol-level approaches, while technically feasible and insightful from a research perspective, require further advances in model efficiency and accelerator hardware before they can be considered for large-scale commercial systems.

Chapter 8

Conclusion and Future Work

This chapter provides an overall conclusion and highlights the contributions of this work. Additionally, a section is dedicated to discussing the potential future work to further progress neural network (NN)-based precoding that tackles hardware nonlinearities.

8.1 Conclusion

The contributions and conclusions are discussed based on the research questions introduced in Chapter 1.

Q1 - Which classical precoding techniques exist to mitigate nonlinear power amplifier (PA) distortion and what are their limitations?

This thesis builds on existing work on classical distortion-aware precoding, which has established that nonlinear distortion can, in principle, be spatially shaped and steered away from intended users. However, such approaches typically rely on simplified signal and hardware models, including simplified PA models and single-user assumptions, which limits their applicability to realistic multi-user, multi-carrier systems. As a result, while classical techniques such as the Z3RO precoder provide valuable insight and serve as important baselines, their direct extension to massive multiple input multiple output (MIMO) deployments with practical hardware constraints remains challenging. These limitations motivate the development of alternative approaches that retain the distortion-shaping capability while considering realistic PA models and multi-user, multi-carrier scenarios.

Q2 - Can NN-based precoding be leveraged to allow PAs to operate close to saturation, while maintaining the data rate?

The results of this thesis demonstrate that neural network-based precoding provides a viable mechanism to operate power amplifiers closer to saturation without sacrificing communication performance or for the same saturation point/back off, improve communication performance (data rate). By learning the nonlinear relationship between the channel and precoder matrix, neural networks can perform distortion-aware precoding in scenarios where analytical designs are infeasible. Importantly, this conclusion holds for both single and multi-carrier systems.

Q2.1 - Which neural network architectures are suitable for linear precoding?

A central conclusion of this work is that the neural network's architectural inductive bias is essential for successful neural network-based precoding. By embedding physical properties such as permutation properties into the model design, graph neural networks achieve strong generalization with manageable complexity. This architectural choice proves critical across single-carrier and multi-carrier systems. Without such structure, generic neural networks exhibit poor scalability and limited performance. A key take-away is that neural networks are only effective for precoding when they explicitly respect the structure of the underlying precoding problem.

Q2.2 - What is the impact of varying channel conditions (different training distribution vs. testing distribution) on the performance of NN-based precoders?

NN-based precoders exhibit robustness to moderate variations in channel statistics, with performance degradation primarily affecting worst-case rather than average behavior. Importantly, the results demonstrate that exposure to diverse channel conditions during training significantly improves robustness, highlighting the importance of representative and heterogeneous training data. These findings underline the need for adaptive or continual learning mechanisms in non-stationary wireless environments. This is especially important when model mismatch occurs between the training time assumptions (channel model, PA model, etc.) and the conditions encountered during deployment.

Q2.3 - In a multi-carrier system, where inter-carrier interference occurs due to the nonlinear PA functions, how can NN-based distortion-aware precoding be used/adapted to allow PAs to operate close to saturation while maintaining a high data rate?

It is demonstrated that in multi-carrier systems, distortion-aware precoding must explicitly account for the coupling between subcarriers introduced by nonlinear

power amplifiers. Unlike single-carrier (frequency-flat) settings, where distortion can be mitigated on a per-carrier basis, nonlinear amplification causes spectral regrowth and inter-carrier interference that links the precoding policy across all carriers. Neural network-based precoding provides a flexible framework to address this challenge by jointly learning precoding strategies across carriers. The results show that when such cross-carrier dependencies are incorporated into the training, power amplifiers can be operated closer to saturation while maintaining robust data rates, even under frequency-selective channel conditions.

Q3 - How can NN-based nonlinear precoding architectures be designed such that the number of digital-to-analog converter (DAC) bits can be reduced without degrading the data rate?

The thesis demonstrates that neural network-based nonlinear precoding can compensate for severe quantization effects by shaping quantization distortion spatially. This enables reliable communication with reduced DAC resolution, highlighting the potential of learning-based methods in highly constrained hardware settings. At the same time, the results reveal a fundamental trade-off: nonlinear precoding requires symbol-level inference, which significantly increases computational demands. This places practical limits on its applicability with current compute hardware for NN inference. Consequently, nonlinear neural network-based precoding enables operation with low resolution DACs, but at a substantial computational cost.

Q4 - What is the potential power consumption reduction of the developped NN-based precoders?

This question is split up into two subquestions, covering the power consumption for the PA (Q4.1) and DAC (Q4.2) cases, both questions are jointly discussed. It is shown that the power reduction enabled through NN-based precoding can only be meaningfully assessed at the system level, where digital processing costs and analog hardware savings are jointly considered. By enabling distortion-aware operation, NN-based precoders allow transmit hardware to be operated in a more energy-efficient manner, leading to substantial power consumption reductions in analog components such as PAs and, to a lesser extent, DACs. In scenarios where precoding can be performed at the channel coherence time, the additional inference power remains manageable and is outweighed by the resulting hardware savings. This is the case for the schemes targeting the PA. In contrast, symbol-level nonlinear precoding introduces a much higher computational burden, which currently limits its practical power reduction despite its potential for aggressive hardware simplification for DACs. A future perspective that could overcome this limitation is to further propagate this bit width reduction into the base

band processing, which could translate into lower power usage in multipliers, adders, registers, memory, and interconnects. These secondary power/complexity reduction effects could motivate further exploration of this work. However, the current findings in this thesis highlight that NN-based precoding is not universally power-saving by default, but can deliver net power reductions when its computational complexity is manageable. More broadly, the results emphasize the importance of applying these NN-based precoders in appropriate system settings, which allow for net power reductions.

8.2 Future Work

This thesis has demonstrated that NN-based precoding is a promising tool for addressing hardware nonlinearities in massive MIMO systems and for providing power reductions at the transmitter. At the same time, the results reveal several open research challenges that must be addressed before such techniques can be widely adopted in practical deployments.

A first important research direction concerns regulatory compliance and spectral behavior as highlighted in Chapter 7. While distortion-aware precoding actively reshapes nonlinear distortion, it does not eliminate it, in fact, operating PAs closer to saturation increases the overall nonlinear distortion. As a result, out-of-band emissions and spectral masks remain critical constraints. Future work should therefore explore learning-based precoding strategies that explicitly incorporate regulatory metrics into the optimization process, enabling solutions that jointly balance communication performance, power reduction, and spectral containment. More broadly, this raises the question of how learning-based transmitters can be designed to be regulation-aware by construction, noting that defining and formalizing the relevant regulatory metrics into implementable optimization constraints can be challenging. Alternatively, a longer-term perspective is to move towards a system in which all users in the main and neighboring spectral bands are identified, and in which distortion is controlled such that no user is contaminated by in-band or out-of-band emissions. This would relax the need for strict spectral masks in all directions, and instead focus the emission constraints on the directions of interest, i.e., the user directions.

A second key challenge is scalability and computational efficiency. As wireless systems evolve toward larger antenna arrays, wider bandwidths, and higher user densities, the complexity of neural network inference may become a limiting factor. Future research should aim to better understand how model size and inference cost scale with system dimensions (transmit antennas, users, carriers), and develop

strategies that mitigate this growth. This includes architectural improvements and model compression techniques such as pruning the NN, reducing the precision of the model weights, knowledge distillation, etc. Finally, higher-level schemes could be explored, such as grouping sufficiently similar sub-carriers to reduce the number of required precoding matrices. Such efforts are essential to ensure that the power savings enabled in the analog domain are not offset by excessive digital processing costs.

A third direction relates to robust deployment and long-term operation. Neural network-based precoders rely on training data that reflects the operating environment, yet real-world wireless channels are inherently non-stationary. Changes in propagation conditions, hardware characteristics, or network topology can gradually degrade performance. Future work should therefore investigate continual learning and adaptation mechanisms that allow precoders to remain effective over time, as well as training strategies that explicitly account for channel estimation errors and model mismatch.

Finally, a further direction is to extend distortion-aware graph neural network precoding to dense multi-user and cell-free deployments. The current thesis has focused on critical cases in terms of non-linear distortion in massive MIMO systems, i.e., scenarios with few users and sufficiently flat fading channels where distortion coherently combines at the user location. However, in dense multi-user and/or cell-free scenarios, the frequency-selectivity of the channel and user spatial correlation will determine whether distortion adds constructively or averages out. As such, in these scenarios, adaptive techniques could be designed that perform distortion-aware precoding when beneficial, and fall back to simpler schemes when channels are close to independently and identically distributed (i.i.d.) in space and frequency, and only limited gains are expected.

More broadly, this thesis contributes to the ongoing paradigm shift in wireless system design, moving from optimizing purely for maximum performance to jointly considering energy reduction. Taken together, the results of this thesis show that hardware nonlinearities should no longer be treated as an unavoidable impairment that is compensated through conservative operating regimes. Instead, they can be actively incorporated into the signal design through learning-based precoding, enabling new trade-offs between performance, energy reduction, and hardware complexity. Bridging the gap between theoretical gains and practical, scalable, and energy reducing implementations remains a central challenge, and a defining opportunity for next-generation wireless systems.

Bibliography

[1] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, O. Blume, and A. Fehske, "How much energy is needed to run a wireless network?," *IEEE wireless communications*, vol. 18, no. 5, pp. 40–49, 2011.

[2] L. Golard, Y. Agram, F. Rottenberg, F. Quitin, D. Bol, and J. Louveaux, "A Parametric Power Model of Multi-Band Sub-6 GHz Cellular Base Stations Using On-Site Measurements," in *2024 IEEE 35th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pp. 1–7, 2024.

[3] "Ericsson Mobility report November 2021." `https://www.ericsson.com/4ad7e9/assets/local/reports-papers/mobility-report/documents/2021/ericsson-mobility-report-november-2021.pdf`.

[4] "Ericsson Mobility report June 2025." `https://www.ericsson.com/en/reports-and-papers/mobility-report/reports/june-2025`.

[5] C. Freitag, M. Berners-Lee, K. Widdicks, B. Knowles, G. S. Blair, and A. Friday, "The real climate and transformative impact of ICT: A critique of estimates, trends, and regulations," *Patterns*, vol. 2, no. 9, p. 100340, 2021.

[6] O. Edfors, R. Brazil, H. Petautschnig, G. Callebaut, T. Feys, L. V. der Perre, E. G. Larsson, O. Edfors, E. Fitzgerald, L. Liu, J. R. Sanchez, W. Tärneberg, P. Frenger, B. Deutschmann, T. Wilding, and K. Witrisal, "Deliverable 2.1: Initial assessment of architectures and hardware resources for a RadioWeaves infrastructure," Jan. 2022.

[7] C. Andersson, J. Bengtsson, G. Byström, P. Frenger, Y. Jading, and M. Nordenström, "Improving energy performance in 5G networks and beyond," *Ericsson Technology Review*, vol. 2022, no. 8, pp. 2–11, 2022.

[8] T. Bateman, "China's 5G network is causing its carbon emissions to soar." https://www.euronews.com/next/2021/05/28/superfast-but-not-so-clean-china-s-5g-network-is-causing-its-carbon-emissions-to-soar.

[9] L. Belkhir and A. Elmeligi, "Assessing ICT global emissions footprint: Trends to 2040 & recommendations," *Journal of Cleaner Production*, vol. 177, pp. 448–463, Mar. 2018.

[10] European Commission, "The European Green Deal," *COM (2019)*, November 2019.

[11] H. Viswanathan, S. Wesemann, J. Du, and H. Holma, "Energy efficiency in next-generation mobile networks," *White Paper, Nokia Bell Labs*, 2022.

[12] J. Joung, C. K. Ho, and S. Sun, "Tradeoff of spectral and energy efficiencies: Impact of power amplifier on OFDM systems," in *2012 IEEE Global Communications Conference (GLOBECOM)*, pp. 3274–3279, IEEE, 2012.

[13] S. Cui, A. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," *IEEE Transactions on Wireless Communications*, vol. 4, no. 5, pp. 2349–2360, 2005.

[14] B. Murmann, "ADC Performance Survey 1997-2025." [Online]. Available: `https://github.com/bmurmann/ADC-survey`.

[15] A. Shafie, N. Yang, C. Han, J. M. Jornet, M. Juntti, and T. Kürner, "Terahertz Communications for 6G and Beyond Wireless Networks: Challenges, Key Advancements, and Opportunities," *IEEE Network*, vol. 37, no. 3, pp. 162–169, 2023.

[16] ZTE Corporation, "Everybit Going Extremely Green Whitepaper: On the way to Mobile Net Zero," 2024. White Paper.

[17] ZTE Corporation, "Remake Green 5G: Mobile Innovation for Climate Action," 2023. White Paper.

[18] E. Björnson, L. Sanguinetti, H. Wymeersch, J. Hoydis, and T. L. Marzetta, "Massive MIMO is a reality-What is next?," *Digital signal processing*, vol. 94, pp. 3–20, 2019.

[19] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge University Press, 2016.

[20] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency," *Foundations and Trends in Signal Processing*, vol. 11, pp. 154–655, 11 2017.

[21] L. Golard, J. Louveaux, and D. Bol, "Evaluation and projection of 4G and 5G RAN energy footprints: the case of Belgium for 2020–2025," *Annals of Telecommunications*, vol. 78, pp. 313–327, 2023.

[22] J. Malmodin, N. Lövehagen, P. Bergmark, and D. Lundén, "ICT sector electricity consumption and greenhouse gas emissions – 2020 outcome," *Telecommunications Policy*, vol. 48, no. 3, p. 102701, 2024.

[23] K. G and B. P, "Energy Consumption in Data Centres and Broadband Communication Networks in the EU," *Publications Office of the European Union*, no. KJ-NA-31-841-EN-N (online), 2024.

[24] B. Graver, K. Zhang, and D. Rutherford, "CO2 Emissions from Commercial Aviation, 2018," Tech. Rep. 2019-16, The International Council on Clean Transportation, 2019.

[25] International Energy Agency, "Aviation – Analysis." `https://www.iea.org/energy-system/transport/aviation`, 2025. Accessed: 2025-11-17.

[26] T. Feys, G. Callebaut, L. Van der Perre, and F. Rottenberg, "Measurement-Based Validation of Z3RO Precoder to Prevent Nonlinear Amplifier Distortion in Massive MIMO Systems," in *2022 IEEE 95th Vehicular Technology Conference: (VTC2022-Spring)*, pp. 1–5, 2022.

[27] T. Feys, X. Mestre, and F. Rottenberg, "Self-Supervised Learning of Linear Precoders Under Non-Linear PA Distortion for Energy-Efficient Massive MIMO Systems," in *ICC 2023 - IEEE International Conference on Communications*, pp. 6367–6372, 2023.

[28] T. Feys, L. Van der Perre, and F. Rottenberg, "Toward Energy-Efficient Massive MIMO: Graph Neural Network Precoding for Mitigating Non-Linear PA Distortion," *IEEE Transactions on Cognitive Communications and Networking*, vol. 11, no. 1, pp. 184–201, 2025.

[29] T. Feys, L. Van der Perre, G. Callebaut, and F. Rottenberg, "Learning When to Learn: Distortion-Aware GNN Precoding for Multi-Carrier Large Antenna Systems with Adaptive Precoder Selection," in *ICC 2026 - IEEE International Conference on Communications*, 2026. (accepted).

[30] T. Feys, L. Van der Perre, and F. Rottenberg, "Self-Supervised Learning of Graph Neural Networks for Precoding in 1-bit Quantized Large Array Systems," in *2025 IEEE International Conference on Machine Learning for Communication and Networking (ICMLCN)*, pp. 1–6, 2025.

[31] T. Feys, L. Van der Perre, and F. Rottenberg, "Learning to Quantize and Precode in Massive MIMO Systems for Energy Reduction: a Graph Neural Network Approach," *IEEE Journal of Selected Topics in Signal Processing*, pp. 1–15, 2025.

[32] S. Cripps, *RF Power Amplifiers for Wireless Communications*. Artech House Microwave Library, Norwood: Artech House, 2006.

[33] E. G. Larsson and L. Van der Perre, "Out-of-Band Radiation From Antenna Arrays Clarified," *IEEE wireless communications letters*, vol. 7, no. 4, pp. 610–613, 2018.

[34] C. Mollen, U. Gustavsson, T. Eriksson, and E. G. Larsson, "Spatial Characteristics of Distortion Radiated From Antenna Arrays With Transceiver Nonlinearities," *IEEE transactions on wireless communications*, vol. 17, no. 10, pp. 6663–6679, 2018.

[35] C. Studer and E. G. Larsson, "PAR-Aware Large-Scale Multi-User MIMO-OFDM Downlink," *IEEE journal on selected areas in communications*, vol. 31, no. 2, pp. 303–313, 2013.

[36] S. R. Aghdam, S. Jacobsson, U. Gustavsson, G. Durisi, C. Studer, and T. Eriksson, "Distortion-Aware Linear Precoding for Massive MIMO Downlink Systems with Nonlinear Power Amplifiers." `https://arxiv.org/abs/2012.13337`, 2021.

[37] I. Iofedov and D. Wulich, "Distortion cancellation precoding for OFDM-SDMA downlink with nonlinear power amplifiers," in *2015 IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pp. 704–709, IEEE, 2015.

[38] R. Zayani, H. Shaiek, and D. Roviras, "Efficient Precoding for Massive MIMO Downlink Under PA Nonlinearities," *IEEE communications letters*, vol. 23, no. 9, pp. 1611–1615, 2019.

[39] F. Rottenberg, G. Callebaut, and L. Van der Perre, "Z3RO Precoder Canceling Nonlinear Power Amplifier Distortion in Large Array Systems," *arXiv preprint arXiv:2110.07891*, 2021.

[40] G. Callebaut, S. Gunnarsson, A. P. Guevara, F. Tufvesson, S. Pollin, L. Van der Perre, and A. J. Johansson, "Massive MIMO goes Sub-GHz: Implementation and Experimental Exploration for LPWANs," in *2020 54th Asilomar Conference on Signals, Systems, and Computers*, pp. 1101–1105, 2020.

[41] C. Rapp, "Effects of HPA-Nonlinearity on a 4-DPSK/OFDM-Signal for a Digital Sound Broadcasting System.," in *Second European Conf. on Sat. Comm., 22. - 24.10.91, Liege, Belgium.*, pp. 179–184, 1991. LIDO-Berichtsjahr=1991, pages=6,.

[42] J. J. Bussgang, "Crosscorrelation functions of amplitude-distorted Gaussian signals," *Tech. Rep. 216, Research Lab. Electron*, 1952.

[43] O. T. Demir and E. Bjornson, "The Bussgang Decomposition of Nonlinear Systems: Basic Theory and MIMO Extensions [Lecture Notes]," *IEEE Signal Processing Magazine*, vol. 38, no. 1, pp. 131–136, 2021.

[44] S. R. Aghdam, S. Jacobsson, and T. Eriksson, "Distortion-Aware Linear Precoding for Millimeter-Wave Multiuser MISO Downlink," in *2019 IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1–6, 2019.

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

[46] T. Eriksson, W. Cao, and C. Fager, *Nonlinear Effects of Wireless Transceivers*, pp. 1–30. John Wiley & Sons, Ltd, 2019.

[47] J. Zanen, E. Klumperink, and B. Nauta, "Power Efficiency Model for MIMO Transmitters Including Memory Polynomial Digital Predistortion," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 68, no. 4, pp. 1183–1187, 2021.

[48] P. P. Ann and R. Jose, "Comparison of PAPR reduction techniques in OFDM systems," in *2016 International Conference on Communication and Electronics Systems (ICCES)*, pp. 1–5, 2016.

[49] S. K. Mohammed and E. G. Larsson, "Constant-Envelope Multi-User Precoding for Frequency-Selective Massive MIMO Systems," *IEEE Wireless Communications Letters*, vol. 2, no. 5, pp. 547–550, 2013.

[50] C. Tarver, A. Balalsoukas-Slimining, C. Studer, and J. R. Cavallaro, "Virtual DPD Neural Network Predistortion for OFDM-based MU-Massive MIMO," in *2021 55th Asilomar Conference on Signals, Systems, and Computers*, pp. 376–380, 2021.

[51] X. Cheng, R. Zayani, M. Ferecatu, and N. Audebert, "Efficient Autoprecoder-based deep learning for massive MU-MIMO Downlink under PA Non-Linearities," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1039–1044, 2022.

[52] J. Jee, G. Kwon, and H. Park, "Joint Precoding and Power Allocation for Multiuser MIMO System With Nonlinear Power Amplifiers," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 9, pp. 8883–8897, 2021.

[53] B. Liu, F. Rottenberg, and S. Pollin, "Power Allocation for Distributed Massive LoS MIMO with Nonlinear Power Amplifiers," in *2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall)*, pp. 1–5, 2022.

[54] J. Jee, G. Kwon, and H. Park, "Cooperative Beamforming With Nonlinear Power Amplifiers: A Deep Learning Approach for Distributed Networks," *IEEE Transactions on Vehicular Technology*, pp. 1–16, 2022.

[55] F. Rottenberg, G. Callebaut, and L. Van der Perre, "The Z3RO Family of Precoders Cancelling Nonlinear Power Amplification Distortion in Large Array Systems," *IEEE Transactions on Wireless Communications*, vol. 22, no. 3, pp. 2036–2047, 2023.

[56] T. Hwang, C. Yang, G. Wu, S. Li, and G. Ye Li, "OFDM and Its Wireless Applications: A Survey," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 4, pp. 1673–1694, 2009.

[57] T. Schenk, *RF Imperfections in High-rate Wireless Systems: Impact and Digital Compensation*. Dordrecht: Springer Netherlands, 2008.

[58] Nokia, "Realistic power amplifier model for the New Radio evaluation," *3GPP TSG-RAN WG4 Meeting 79, R4-163314*, May 2016.

[59] L. Anttila, A. Brihuega, and M. Valkama, "On Antenna Array Out-of-Band Emissions," *IEEE Wireless Communications Letters*, vol. 8, no. 6, pp. 1653–1656, 2019.

[60] C.-S. Choi *et al.*, "RF impairment models for 60GHz-band SYS/PHY simulation," *Project: IEEE P802.15 Working Group for Wireless Personal Area Networks (WPANs)*, p. 17, 2006.

[61] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," 2014.

[62] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural networks*, vol. 2, no. 5, pp. 359–366, 1989.

[63] N. Keriven and G. Peyré, "Universal Invariant and Equivariant Graph Neural Networks," 2019.

[64] B. Zhao, J. Guo, and C. Yang, "Understanding the Performance of Learning Precoding Policy with GNN and CNNs," 2024.

[65] P. W. Battaglia *et al.*, "Relational inductive biases, deep learning, and graph networks," 2018. 10.48550/ARXIV.1806.01261.

[66] M. Anthony and P. L. Bartlett, *Neural Network Learning: Theoretical Foundations*. Cambridge University Press, 1999.

[67] B. Zhao, J. Guo, and C. Yang, "Learning Precoding Policy: CNN or GNN?," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1027–1032, 2022.

[68] W. L. Hamilton, "Graph Representation Learning," *Synthesis Lectures on Artificial Intelligence and Machine Learning*, vol. 14, no. 3, pp. 1–159, 2020.

[69] C. Silvano, D. Ielmini, F. Ferrandi, L. Fiorin, S. Curzel, L. Benini, F. Conti, A. Garofalo, C. Zambelli, E. Calore, S. F. Schifano, M. Palesi, G. Ascia, D. Patti, S. Perri, N. Petra, D. D. Caro, L. Lavagno, T. Urso, V. Cardellini, G. C. Cardarilli, and R. Birke, "A Survey on Deep Learning Hardware Accelerators for Heterogeneous HPC Platforms," 2023.

[70] F. Piazza, A. Uncini, and M. Zenobi, "Neural networks with digital LUT activation functions," in *Proceedings of 1993 International Conference on Neural Networks (IJCNN-93-Nagoya, Japan)*, vol. 2, pp. 1401–1404 vol.2, 1993.

[71] R. Hunger, *Floating Point Operations in Matrix-vector Calculus*. Munich University of Technology, Inst. for Circuit Theory and Signal Processing, 2005.

[72] L. Baischer, M. Wess, and N. TaheriNejad, "Learning on Hardware: A Tutorial on Neural Network Accelerators and Co-Processors," 2021.

[73] A. He, S. Srikanteswara, K. K. Bae, T. R. Newman, J. H. Reed, W. H. Tranter, M. Sajadieh, and M. Verhelst, "Power Consumption Minimization for MIMO Systems — A Cognitive Radio Approach," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 2, pp. 469–479, 2011.

[74] Ericsson, "Antenna Integrated Radio Unit Description: AIR 6488." https://www.1com.net/wp-content/uploads/2019/09/sales@1com.com-Ericsson-AIR-6488-Integrated-Radio-Unit-Datasheet.pdf, jun 2019.

[75] S. Malkowsky, J. Vieira, K. Nieman, N. Kundargi, I. Wong, V. Öwall, O. Edfors, F. Tufvesson, and L. Liu, "Implementation of Low-Latency Signal Processing and Data Shuffling for TDD Massive MIMO Systems," in *2016 IEEE International Workshop on Signal Processing Systems (SiPS)*, pp. 260–265, 2016.

[76] S. Malkowsky, J. Vieira, L. Liu, P. Harris, K. Nieman, N. Kundargi, I. C. Wong, F. Tufvesson, V. Öwall, and O. Edfors, "The World's First Real-Time Testbed for Massive MIMO: Design, Implementation, and Validation," *IEEE Access*, vol. 5, pp. 9073–9088, 2017.

[77] F. Zaruba, F. Schuiki, and L. Benini, "Manticore: A 4096-Core RISC-V Chiplet Architecture for Ultraefficient Floating-Point Computing," *IEEE Micro*, vol. 41, pp. 36–42, Mar. 2021.

[78] M. Sugiyama, M. Krauledat, and K.-R. Müller, "Covariate Shift Adaptation by Importance Weighted Cross Validation," *Journal of Machine Learning Research*, vol. 8, no. 35, pp. 985–1005, 2007.

[79] J. Quionero-Candela, M. Sugiyama, A. Schwaighofer, and N. D. Lawrence, *Dataset Shift in Machine Learning*. The MIT Press, 2009.

[80] D. Li, Y. Yang, Y.-Z. Song, and T. M. Hospedales, "Learning to generalize: meta-learning for domain generalization," in *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence*, AAAI'18/IAAI'18/EAAI'18, AAAI Press, 2018.

[81] A. H. Karkan, A. Ibrahim, J.-F. Frigon, and F. Leduc-Primeau, "A Low-Complexity Plug-and-Play Deep Learning Model for Massive MIMO Precoding Across Sites," 2025.

[82] T. Miao, T. Feys, G. Callebaut, J. Van Mulders, E. Peschiera, M. A. Rahman, and F. Rottenberg, "GNN-based Precoder Design and Fine-tuning for Cell-free Massive MIMO with Real-world CSI," *Proceedings of the 2025 Symposium on Information Theory and Signal Processing in the Benelux*, 2025-05-19.

[83] T. Miao, T. Feys, G. Callebaut, J. Van Mulders, M. A. Rahman, and F. Rottenberg, "Testbed Evaluation of AI-based Precoding in Distributed MIMO Systems," in *2026 IEEE Wireless Communications and Networking Conference (WCNC)*, 2026. (accepted).

[84] D. Delabie, B. Cox, L. De Strycker, and L. Van der Perre, "Techtile: a Flexible Testbed for Distributed Acoustic Indoor Positioning and Sensing," in *2022 IEEE Sensors Applications Symposium (SAS)*, pp. 1–6, 2022.

[85] A. M. Elbir, K. V. Mishra, M. R. B. Shankar, and B. Ottersten, "A Family of Deep Learning Architectures for Channel Estimation and Hybrid Beamforming in Multi-Carrier mm-Wave Massive MIMO," *IEEE Transactions on Cognitive Communications and Networking*, vol. 8, no. 2, pp. 642–656, 2022.

[86] S. Jacobsson, G. Durisi, M. Coldrey, and C. Studer, "Linear Precoding With Low-Resolution DACs for Massive MU-MIMO-OFDM Downlink," *IEEE Transactions on Wireless Communications*, vol. 18, no. 3, pp. 1595–1609, 2019.

[87] L. H. Crockett, D. Northcote, and R. W. Steward, eds., *Software defined radio with Zynq Ultrascale+ RFSoC*. Strathclyde Academic Media, 1 ed., 2023.

[88] A. van den Bosch, M. Steyaert, and W. Sansen, "SFDR-bandwidth limitations for high speed high resolution current steering CMOS D/A converters," in *ICECS'99. Proceedings of ICECS '99. 6th IEEE International Conference on Electronics, Circuits and Systems (Cat. No.99EX357)*, vol. 3, pp. 1193–1196 vol.3, 1999.

[89] L. Fan, S. Jin, C.-K. Wen, and H. Zhang, "Uplink Achievable Rate for Massive MIMO Systems With Low-Resolution ADC," *IEEE Communications Letters*, vol. 19, no. 12, pp. 2186–2189, 2015.

[90] O. Orhan, E. Erkip, and S. Rangan, "Low power analog-to-digital conversion in millimeter wave systems: Impact of resolution and bandwidth on performance," in *2015 Information Theory and Applications Workshop (ITA)*, pp. 191–198, 2015.

[91] H. Pirzadeh and A. L. Swindlehurst, "Spectral Efficiency of Mixed-ADC Massive MIMO," *IEEE Transactions on Signal Processing*, vol. 66, no. 13, pp. 3599–3613, 2018.

[92] J. Zhang, L. Dai, Z. He, S. Jin, and X. Li, "Performance Analysis of Mixed-ADC Massive MIMO Systems Over Rician Fading Channels," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 6, pp. 1327–1338, 2017.

[93] J. Chae and J. H. Cho, "An ADC-Aware Receiver Design for Multi-User MIMO Underlay System With Strong Cyclostationary Legacy Signal," *IEEE Access*, vol. 11, pp. 119788–119801, 2023.

[94] A. K. Saxena, I. Fijalkow, and A. L. Swindlehurst, "On one-bit quantized ZF precoding for the multiuser massive MIMO downlink," in *2016 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pp. 1–5, 2016.

[95] O. B. Usman, H. Jedda, A. Mezghani, and J. A. Nossek, "MMSE precoder for massive MIMO using 1-bit quantization," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3381–3385, 2016.

[96] S. Jacobsson, G. Durisi, M. Coldrey, T. Goldstein, and C. Studer, "Quantized Precoding for Massive MU-MIMO," *IEEE Transactions on Communication*, vol. 65, pp. 4670–4684, Nov. 2017.

[97] S. Jacobsson, G. Durisi, M. Coldrey, T. Goldstein, and C. Studer, "Nonlinear 1-bit precoding for massive MU-MIMO with higher-order modulation," in *2016 50th Asilomar Conference on Signals, Systems and Computers*, pp. 763–767, 2016.

[98] A. Mezghani, R. Ghiat, and J. A. Nossek, "Transmit processing with low resolution D/A-converters," in *2009 16th IEEE International Conference on Electronics, Circuits and Systems - (ICECS 2009)*, pp. 683–686, 2009.

[99] J. Choi, J. Park, and N. Lee, "Energy Efficiency Maximization Precoding for Quantized Massive MIMO Systems," *IEEE Transactions on Wireless Communications*, vol. 21, no. 9, pp. 6803–6817, 2022.

[100] E. Björnson, L. Sanguinetti, and J. Hoydis, "Hardware Distortion Correlation Has Negligible Impact on UL Massive MIMO Spectral Efficiency," *IEEE Transactions on Communication*, vol. 67, no. 2, pp. 1085–1098, 2019.

[101] L. N. Ribeiro, S. Schwarz, M. Rupp, and A. L. F. de Almeida, "Energy Efficiency of mmWave Massive MIMO Precoding With Low-Resolution DACs," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 2, pp. 298–312, 2018.

[102] H. Huang, W. Xia, J. Xiong, J. Yang, G. Zheng, and X. Zhu, "Unsupervised Learning-Based Fast Beamforming Design for Downlink MIMO," *IEEE Access*, vol. 7, pp. 7599–7605, 2019.

[103] W. Xia, G. Zheng, K.-K. Wong, and H. Zhu, "Model-Driven Beamforming Neural Networks," *IEEE Wireless Communications*, vol. 27, no. 1, pp. 68–75, 2020.

[104] T. Feys, X. Mestre, E. Peschiera, and F. Rottenberg, "Deep Unfolding for Fast Linear Massive MIMO Precoders under a PA Consumption Model," in *2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring)*, pp. 1–5, 2023.

[105] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[106] A. Balatsoukas-Stimming, O. Castañeda, S. Jacobsson, G. Durisi, and C. Studer, "Neural-Network Optimized 1-bit Precoding for Massive MU-MIMO," in *2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 1–5, 2019.

[107] O. Castañeda, S. Jacobsson, G. Durisi, M. Coldrey, T. Goldstein, and C. Studer, "1-bit Massive MU-MIMO Precoding in VLSI," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 7, no. 4, pp. 508–522, 2017.

[108] M. Hossienzadeh, H. Aghaeinia, and M. Kazemi, "Deep Learning Based Interference Exploitation in 1-Bit Massive MIMO Precoding," *IEEE Access*, vol. 11, pp. 17096–17103, 2023.

[109] I. Bello, H. Pham, Q. V. Le, M. Norouzi, and S. Bengio, "Neural Combinatorial Optimization with Reinforcement Learning," 2017.

[110] A. I. Garmendia, J. Ceberio, and A. Mendiburu, "Neural Combinatorial Optimization: a New Player in the Field," 2022.

[111] Y. Bengio, A. Lodi, and A. Prouvost, "Machine learning for combinatorial optimization: A methodological tour d'horizon," *European Journal of Operational Research*, vol. 290, no. 2, pp. 405–421, 2021.

[112] R. Wang, L. Shen, Y. Chen, X. Yang, D. Tao, and J. Yan, "Towards One-shot Neural Combinatorial Solvers: Theoretical and Empirical Notes on the Cardinality-Constrained Case," in *The Eleventh International Conference on Learning Representations*, 2023.

[113] I. A. M. Huijben, W. Kool, M. B. Paulus, and R. J. G. van Sloun, "A Review of the Gumbel-max Trick and its Extensions for Discrete Stochasticity in Machine Learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 2, pp. 1353–1371, 2023.

[114] A. K. Fletcher, S. Rangan, V. K. Goyal, and K. Ramchandran, "Robust Predictive Quantization: Analysis and Design Via Convex Optimization," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 618–632, 2007.

[115] A. Gersho and R. Gray, *Vector Quantization and Signal Compression*. The Springer International Series in Engineering and Computer Science, Springer US, 1991.

[116] J. Max, "Quantizing for minimum distortion," *IRE Transactions on Information Theory*, vol. 6, no. 1, pp. 7–12, 1960.

[117] E. Jang, S. Gu, and B. Poole, "Categorical Reparameterization with Gumbel-Softmax," 2017.

[118] K. Guo, W. Li, Z. Zhu, S. Zeng, T. Xie, S. HA, Y. Xie, P. Debacker, M. Verhelst, and Y. Wang, "Neural Network Accelerator Comparison." https://nicsefc.ee.tsinghua.edu.cn/networkone.html.

[119] Z.-S. Fu, Y.-C. Lee, A. Park, and C.-H. Yang, "A 40-nm 646.6TOPS/W Sparsity-Scaling DNN Processor for On-Device Training," in *2022 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits)*, pp. 40–41, 2022.

[120] L. Wang, X. Zhang, H. Su, and J. Zhu, "A Comprehensive Survey of Continual Learning: Theory, Method and Application," 2024.

[121] Juan Montojo, "Overview of AI/ML related work in 3GPP." ETSI AI Conference 2025.

[122] European Commission, "A European strategy for data," *COM(2020) 66 final*, 2020.

[123] Government of Flanders, "Flemish Reform Programme." `https://www.flandersineu.be/sites/default/files/atoms/files/FlemishReformProgramme2020_0.pdf`, 2020.

[124] E. Ekudden, "Breaking the energy curve: how to roll out 5G without increasing energy consumption."

[125] "Brussels Airport innovates with 5G network." https://www.brusselsairport.be/brusselsairportnews/en/december-2019/brussels-airport-innovates-with-private-5g-network.

[126] "Test with 5G in the port." https://www.portofantwerp.com/en/news/test-5g-port.

[127] S. Ali, W. Saad, N. Rajatheva, K. Chang, D. Steinbach, B. Sliwa, C. Wietfeld, K. Mei, H. Shiri, H.-J. Zepernick, T. M. C. Chu, I. Ahmad, J. Huusko, J. Suutala, S. Bhadauria, V. Bhatia, R. Mitra, S. Amuru, R. Abbas, B. Shao, M. Capobianco, G. Yu, M. Claes, T. Karvonen, M. Chen, M. Girnyk, and H. Malik, "6G White Paper on Machine Learning in Wireless Communication Networks," 2020.

[128] T. O'Shea and J. Hoydis, "An Introduction to Deep Learning for the Physical Layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, 2017.

[129] N. Ye, S. Miao, J. Pan, Q. Ouyang, X. Li, and X. Hou, "Artificial Intelligence for Wireless Physical-Layer Technologies (AI4PHY): A Comprehensive Survey," *IEEE Transactions on Cognitive Communications and Networking*, vol. 10, no. 3, pp. 729–755, 2024.

[130] N. Van Huynh, J. Wang, H. Du, D. T. Hoang, D. Niyato, D. N. Nguyen, D. I. Kim, and K. B. Letaief, "Generative AI for Physical Layer Communications: A Survey," *IEEE Transactions on Cognitive Communications and Networking*, vol. 10, no. 3, pp. 706–728, 2024.

[131] SNS Joint Undertaking Reliable Software Networks Working Group, "The AI/ML Landscape for Smart Networks and Services," white paper, Smart Networks and Services Joint Undertaking (SNS JU), January 2026. Explores the role of AI and Machine Learning in advancing smart networks and services, clarifying terminology, summarising standardisation efforts, and describing innovation pathways for AI-native systems in the context of 6G networks.

[132] K. Trichias, M. Christopoulou, P. Porambage, C. Benzaid, V. Vassiliou, D. Artuñedo Guillén, D. Tsolkas, M. Avgeris, C. Papagianni, A. Jain, D. Kilper, and T. Chen, "AI/ML as a Key Enabler of 6G Networks: Methodology, Approach and AI-Mechanisms in SNS JU ," Jan. 2025.

[133] K. Dyer, "6G meeting settles on same old air interface." `https://the-mobile-network.com/2025/08/6g-meeting-settles-on-same-old-air-interface/`, August 28 2025. The Mobile Network.

[134] Z. Cui, P. Zhang, and S. Pollin, "6G Wireless Communications in 7–24 GHz Band: Opportunities, Techniques, and Challenges," in *2025 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, pp. 1–8, 2025.

[135] G. Xu, Y. Nam, S. Lee, A. Tehrani, D. Y. Lee, A. Prasad, C. Tarver, and M. Tonnemacher, "Upper Mid-Band Spectrum for 6G: Opportunities

and Key Enablers." https://research.samsung.com/blog/Upper-Mid-Band-Spectrum-for-6G-Opportunities-and-Key-Enablers, 2025. Samsung Research blog.

[136] D. Y. Lie, J. C. Mayeda, and J. Lopez, "RF Hardware Design for 5G mm-Wave and 6G Revolutions: Challenges and Opportunities," in *2021 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*, pp. 1–2, 2021.

[137] European Telecommunications Standards Institute (ETSI), "ETSI TS 138 104 V16.7.0: 5G; NR; Base Station (BS) radio transmission and reception (3GPP TS 38.104 version 16.7.0 Release 16)," Technical Specification TS 138 104 V16.7.0, ETSI, Apr 2021.

[138] European Parliament and Council of the European Union, "Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)." Regulation, Official Journal of the European Union, jun 2024. CELEX: 32024R1689.

[139] "AI Act | Shaping Europe's digital future." https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai, 2026.

[140] F. Helbig and C. Mennen, "The AI Act Explained." Forvis Mazars & Pellicaan Advocaten, 2025.

[141] "Directive 2014/53/EU of the European Parliament and of the Council on the harmonisation of the laws of the Member States relating to the making available on the market of radio equipment." Official Journal of the European Union, 2014. Articles 3.1 and 3.2 define safety, electromagnetic compatibility, and efficient use of radio spectrum.

[142] European Parliament and Council of the European Union, "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)." Official Journal of the European Union, L 119, 4 May 2016, pp. 1–88, 2016.

[143] D. Delabie, B. Cox, T. Feys, L. Van der Perre, and L. De Strycker, "Position Dependent Anchor Selection for Scalable Hybrid RF-Acoustic Indoor

Positioning," in *2024 14th International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1–6, 2024.

[144] D. R. Kumar, T. Feys, C. Antón-Haro, F. Rottenberg, and X. Mestre, "Graph Neural Networks for Hybrid Beamforming in MIMO Rate Splitting Multiple Access," in *2025 IEEE International Conference on Machine Learning for Communication and Networking (ICMLCN)*, pp. 1–6, 2025.

[145] D. Delabie, T. Feys, C. Buyle, B. Cox, L. Van der Perre, and L. De Strycker, "Echoes of Accuracy: Enhancing Ultrasonic Indoor Positioning for Energy-Neutral Devices With Neural Network Approaches," *IEEE Journal of Indoor and Seamless Positioning and Navigation*, vol. 3, pp. 227–244, 2025.

[146] R. G. Alavicheh, T. Feys, M. A. Rahman, and F. Rottenberg, "Leveraging Power Amplifier Distortion for Physical Layer Security," in *2025 IEEE 36th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pp. 1–6, 2025.

Statement on the use of Generative AI

I did use the generative AI assistance tools: Perplexity, ChatGPT, and Gemini. Generative AI was used for language assistance and generating/debugging TikZ and Python plotting code.
The text, code, and images in this thesis are my own (unless otherwise specified). Generative AI has only been used in accordance with the KU Leuven guidelines and appropriate references have been added. I have reviewed and edited the content as needed and I take full responsibility for the content of the thesis.

List of Publications

Articles in Internationally Reviewed Academic Journals

T. Feys, L. Van der Perre, and F. Rottenberg, "Toward Energy-Efficient Massive MIMO: Graph Neural Network Precoding for Mitigating Non-Linear PA Distortion," *IEEE Transactions on Cognitive Communications and Networking*, vol. 11, no. 1, pp. 184–201, 2025 [28]

T. Feys, L. Van der Perre, and F. Rottenberg, "Learning to Quantize and Precode in Massive MIMO Systems for Energy Reduction: a Graph Neural Network Approach," *IEEE Journal of Selected Topics in Signal Processing*, pp. 1–15, 2025 [31]

Articles in International Conferences and Symposia

T. Feys, G. Callebaut, L. Van der Perre, and F. Rottenberg, "Measurement-Based Validation of Z3RO Precoder to Prevent Nonlinear Amplifier Distortion in Massive MIMO Systems," in *2022 IEEE 95th Vehicular Technology Conference: (VTC2022-Spring)*, pp. 1–5, 2022 [26]

T. Feys, X. Mestre, and F. Rottenberg, "Self-Supervised Learning of Linear Precoders Under Non-Linear PA Distortion for Energy-Efficient Massive MIMO Systems," in *ICC 2023 - IEEE International Conference on Communications*, pp. 6367–6372, 2023 [27]

T. Feys, X. Mestre, E. Peschiera, and F. Rottenberg, "Deep Unfolding for Fast Linear Massive MIMO Precoders under a PA Consumption Model," in *2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring)*, pp. 1–5, 2023 [104]

T. Feys, L. Van der Perre, and F. Rottenberg, "Self-Supervised Learning of Graph Neural Networks for Precoding in 1-bit Quantized Large Array Systems," in *2025 IEEE International Conference on Machine Learning for Communication and Networking (ICMLCN)*, pp. 1–6, 2025 [30]

T. Feys, L. Van der Perre, G. Callebaut, and F. Rottenberg, "Learning When to Learn: Distortion-Aware GNN Precoding for Multi-Carrier Large Antenna Systems with Adaptive Precoder Selection," in *ICC 2026 - IEEE International Conference on Communications*, 2026. (accepted) [29]

Non-first-author papers

D. Delabie, B. Cox, **T. Feys**, L. Van der Perre, and L. De Strycker, "Position Dependent Anchor Selection for Scalable Hybrid RF-Acoustic Indoor Positioning," in *2024 14th International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1–6, 2024 [143]

D. R. Kumar, **T. Feys**, C. Antón-Haro, F. Rottenberg, and X. Mestre, "Graph Neural Networks for Hybrid Beamforming in MIMO Rate Splitting Multiple Access," in *2025 IEEE International Conference on Machine Learning for Communication and Networking (ICMLCN)*, pp. 1–6, 2025 [144]

D. Delabie, **T. Feys**, C. Buyle, B. Cox, L. Van der Perre, and L. De Strycker, "Echoes of Accuracy: Enhancing Ultrasonic Indoor Positioning for Energy-Neutral Devices With Neural Network Approaches," *IEEE Journal of Indoor and Seamless Positioning and Navigation*, vol. 3, pp. 227–244, 2025 [145]

R. G. Alavicheh, **T. Feys**, M. A. Rahman, and F. Rottenberg, "Leveraging Power Amplifier Distortion for Physical Layer Security," in *2025 IEEE 36th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pp. 1–6, 2025 [146]

T. Miao, **T. Feys**, G. Callebaut, J. Van Mulders, M. A. Rahman, and F. Rottenberg, "Testbed Evaluation of AI-based Precoding in Distributed MIMO Systems," in *2026 IEEE Wireless Communications and Networking Conference (WCNC)*, 2026. (accepted) [83]

T. Miao, **T. Feys**, G. Callebaut, J. Van Mulders, E. Peschiera, M. A. Rahman, and F. Rottenberg, "GNN-based Precoder Design and Fine-tuning for Cell-free Massive MIMO with Real-world CSI," *Proceedings of the 2025 Symposium on Information Theory and Signal Processing in the Benelux*, 2025-05-19 [82]

FACULTY OF ENGINEERING TECHNOLOGY
DEPARTMENT OF ELECTRICAL ENGINEERING
Gebroeders De smetstraat 1
B-9000 Ghent
thomas.feys@kuleuven.be
https://iiw.kuleuven.be/onderzoek/dramco