UPPSALA
UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 1642

Channel Estimation and Prediction for 5G Applications

RIKKE APELFRÖJD

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2018

ISSN 1651-6214
ISBN 978-91-513-0263-8
urn:nbn:se:uu:diva-344270

Dissertation presented at Uppsala University to be publicly examined in Häggsalen, Å10132, Lägerhyddsvägen 1, Uppsala, Friday, 27 April 2018 at 10:00 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Docent Emil Björnson (Linköpings Universitet).

Abstract
Apelfröjd, R. 2018. Channel Estimation and Prediction for 5G Applications. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology* 1642. 116 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-0263-8.

Accurate channel state information (CSI) is important for many candidate techniques of future wireless communication systems. However, acquiring CSI can sometimes be difficult, especially if the user equipment is mobile in which case the future channel realisations must be estimated/predicted. In realistic settings the predictability of radio channels is limited due to measurement noise, limited model orders and since the fading statistics must be modelled based on a set of limited and noisy training data.

In this thesis, the limits of predictability for the radio channel are investigated. Results show that the predictability is limited primarily due to limitations in the training data, while the model order provides a second order limitation effect and the measurement noise comes in as a third order effect.

Then, a Kalman-based linear filter is studied for potential 5G technologies:

C*oherent coordinated multipoint joint transmission*, where channel predictions and the covariance matrix of the prediction error are used to design a robust linear precoder, evaluated in a three base station system. Results show that prediction improves the CSI for the pedestrian users such that system delays of 10 ms are acceptable. The use of the covariance matrix is important for difficult user groups, but of less importance with a simple user grouping system proposed.

Massive multiple-input multiple-output (MIMO) in frequency division duplex (FDD) systems were a reduced, suboptimal, Kalman filter is suggested to estimate channels based on non-orthogonal pilots. By introducing a fixed grid of beams, the system generates sparsity in the channel vectors seen by each user, which then estimates its most relevant channels based on unique pilot codes for each beam. Results show that there is a 5 dB loss compared to orthogonal pilots.

Downlink time division duplex (TDD) channels are estimated based on uplink pilots. By using a predictor antenna, which scouts the channel in advance, the desired downlink channel can be estimated using pilot-based estimates of the channels before and after it (in space). Results indicate that, with the help of Kalman smoothing, predictor antennas can enable accurate CSI for TDD downlinks at vehicular velocities of 80 km/h.

Keywords: Channel estimation, Channel prediction, Channel smoothing, Linear estimation, Kalman filter, Massive MIMO, Coordinated Multipoint transmission, Robust precoding, Predictor antennas, Limits of predictability, Long range predictions

Rikke Apelfröjd, Department of Engineering Sciences, Signals and Systems Group, Box 528, Uppsala University, SE-75120 Uppsala, Sweden.

ISSN 1651-6214
ISBN 978-91-513-0263-8
urn:nbn:se:uu:diva-344270 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-344270)

Till Kasper och Alexander

Sammanfattning

Användningsområdena för trådlös kommunikation ökar ständigt. Applikationer såsom olika streamings-tjänster, arbete mot servrar och så kallade molntjänster gör att fler och fler användare av det trådlösa nätverket önskar ständig uppkoppling och ofta med höga datatakter, oavsett om de är på kontoret, på bussen eller till och med ute på promenad. För att kunna tillgodose användarnas önskemål kommer framtida 5G-system med stor sannoliket att utgöras av en verktygslåda där många olika tekniker finns tillgängliga för att användas av systemet. Två kandidater som har föreslagits för att öka såväl spektraleffektiviteten som täckningen hos ett kommunikationssystem är så kallad massiv MIMO (eng. Multiple-Input-Multiple-Output) och koordinerad multipunktstransmission.

Massiv MIMO bygger på att en basstation med ett mycket stort antal antenner använder dessa för att rikta signalen som är avsedd för an specifik användare mot just denna användare. Denna teknik gör det möjligt att serva ett stort antal användare inom samma resurser, eftersom basstationen har möjlighet att rikta inte bara en, utan ett mycket stort antal signaler (upp till lika många som basstationen har antenner), på samma gång. Genom att serva många användare på samma gång utnyttjas det tillgängliga radiospektrumet bättre, man säger att spektraleffektiviteten ökar.

En av de faktorer som begränsar hur mycket data man kan sända över trådlösa radiokanaler är interferens, störsignaler från andra källor. Ett trådlöst kommunikationssystem är ofta indelat i celler vars gränser bestäms av vilken basstation som har starkast signal i området. Just vid gränserna till dessa celler kan störsignaler från andra basstationer vara extra starka, och därmed försämra täckningen i området kring cellgränsen. Koordinerad multipunktssändning är ett sätt att minska störsignalerna vid cellgränserna och, i bästa fall, förvandla den energi som orsakar störningarna till nyttoenergi. Grundtanken här är att flera basstationer bildar sammarbetskluster. Inom ett kluster delar basstationer information om t.ex. den data som ska skickas till de användare som befinner sig i klustret och information om radiokanalerna till de olika användarna. Genom att koordinera sig kan basstationerna serva alla användare gemensamt.

För bägge dessa två tekniker är det viktigt med kunskap om den så kallade radiokanalen, vilket är en modell av hur radiosignalen påverkas från dess att den lämnat sändaren till dess att den tagits emot av mottagaren.

I denna avhandling används Kalmanfiltrering för att uppskatta radiokanalers egenskaper under olika omständigheter och utvärdera hur dessa skattningar

kan användas för massiv MIMO koordinerad multipunktssändning, och för kommunikation med fordon.

Kalmanfilter är en välkänd metod för att uppskatta och följa hur värdet hos en okänd parameter ändras över tiden utifrån kända mätvärden. I fallet med radiokanaler skickas kända signaler, så kallade piloter, över kanalerna inom vissa givna tids- och frekvensluckor. Piloterna kan vara ortogonala, så att piloter som ska användas för att uppskatta olika radiokanaler skickas på olika tids- och frekvensluckor, eller de kan vara överlappande i vilket fall piloter skickas över olika radiokanaler på samma tids- och frekvensluckor. Medan den tidigare ger möjlighet för mer exakta kanalskattningar gör den senare att man kan använda färre resurser för piloter och därmed frigöra fler resurser till att skicka data.

Om massiv MIMO ska kunna användas i ett system där upplänk (från användare till basstation) och nedlänk (från basstation till användare) separeras i olika frekvensband, s.k. FDD-system som t.ex. är används i dagens 4G-system, så behövs överlappande piloter i nedlänken eftersom antalet antenner hos basstationen är så stort att om alla dessa skulle skicka ortogonala piloter så skulle det bli väldigt lite resurser kvar för att sända data.

Kanalestimering an nedlänkskanaler från en massiv MIMO-antenn i FDD-system studeras i ett av bidragen i avhandlingen. Den lösning som föreslås här bygger på att man, i ett första steg använder s.k. lobformning där olika lober sänder radioeneringen i olika riktningar, vilket får som följd att hos varje användare är det bara ett mindre antal av alla radiokanaler (lober) som är relevanta. Genom att dessutom införa pilotkoder så kommer varje användare att kunna uppskatta just sina egna relevanta radiokanaler. Simuleringsresultaten visar att man på det här sättet kan få radiokanalskattningar som ger nära de prestanda som man kan få med ortogonala piloter och som därmed möjliggör massiv MIMO vinster även i FDD-system.

De största vinsterna hos massiv MIMO bör kunna hämtas om man utnyttjar system där upplänken och nedlänken använder samma frekvensband, men skiljs åt i tid. Sådana system kallas för TDD-system och har fördelen att piloterna som skickas från användarna i upplänken kan användas för att skatta kanalerna i nedlänken, och eftersom antalet användare ofta är färre än antalet antenner hos basstationen i ett massiv MIMO-system kan man använda ortogonala piloter.

En nackdel med TDD-system är att när användare rör på sig kan kanalskattningarna som erhållits baserat på upplänkspiloterna hinna bli gamla innan det är dags att sända data i nedlänken. Detta gäller speciellt vid kommunikation med rörliga fordon. Då kan man behöva prediktera radiokanalen in i framtiden. Samma sak gäller i ett system med koordinerad multipunktssändning eftersom tiden det tar att dela data mellan basstationerna ibland kan vara upp till tiotals millisekunder.

I den här avhandlingen visas, via både uppmätta radiokanaler och teoretiska kanalmodeller, att det finns en gräns för över hur lång sträcka man kan predik-

tera med Kalmanfilter när användaren rör sig igenom en komplicerad radiomiljö. En gräns som ligger runt 0.2-0.3 gånger längden av den våglängd som används för sändning. Hur långt detta motsvarar i tid beror dels på bärvågsfrekvensen och dels på hur snabbt användaren rör sig. Som ett exempel kan nämnas att i ett fall med tidsfördröjningar på 20 ms och en bärvågsfrekvens på 2.65 GHz så är det svårt att prediktera radiokanaler för fordonsburna användare.

Kanalprediktioner via Kalmanfiltrering har utvärderats för långsamma användare (5 km/h) i ett system med koordinerad multipunktstransmission, baserat på tidsserier av uppmätta radiokanaler. För att inte riskera att dåliga prediktioner förstör lösningen föreslås en robust förkodningsalgoritm som inte bara använder sig av prediktionerna utan också av information om hur pålitliga dessa är. Resultaten visar att med hjälp av bägge dessa element, prediktion och robust förkodning, kan man säkra sig vinster vid koordinerad multipunktssändning.

En fördel med den föreslagna robusta förkodningsalgoritmen är att den enkelt kan anpassas för att hantera systembegränsningar i hur mycket data som kan delas mellan basstationerna. Simuleringsresultat visar att detta framförallt är viktigt för att bibehålla så stor del av vinsterna som koordinerat multipunktssändning bidrar med som möjligt vid cellgränserna.

Vidare föreslås en enkel metod för att välja ut vilka användare som ska servas gemensamt på en resurs. Metoden, som bygger på varje basstation schemalägger användare inom sin egen cell baserat på kanalinformation, ökar vinsterna med koordinerat multipunktssändning markant.

När användare färdas i högre hastigheter så fungerar inte längre kanalprediktion som baseras på att man extrapolerar gamla mätningar framåt i tiden. Då kan man istället använda sig av den s.k. prediktionsantennsmetoden. Eftersom höga hastigheter generellt innefattar ett fordon så kan man utnyttja fordonets tak för att där placera två antenner. Om den ena av dessa placeras framför den andra i fordonets färdriktning, så kommer den främre, som kallas prediktionsantennen, att kunna uppskatta radiokanalen innan den bakre antennen, som kallas huvudantennen, upplever samma kanal, och därmed prediktera huvudantennens kanal. Med denna metod kan radiokanaler skattas långt i förväg.

För massiv MIMO i TDD-system innebär prediktionsantenner att huvudantennens nedlänkskanaler kan skattas baserat på såväl tidigare skattningar som skattningar av kanalen i positioner som huvudantennen först senare kommer att nå.

Metoden att uppskatta en parameter baserad på både framtida och tidigare mätningar kallas för glättning. Man strävar efter att uppnå en optimal kombination av brusundertryckning och interpolation och Kalmanalgoritmen är ett optimalt verktyg för detta syfte. Simuleringsresultaten som presenteras i denna avhandling visar att glättning med hjälp av Kalmanfiltrering möjliggör utökade tidsintervall för sändning i nedlänk i ett TDD-system som, vid

höga användarhastigheter, motsvarar en sträcka upp till 0.6-0.7 av utbredningsvåglängden.

List of papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I **Rikke Apelfröjd** "Kalman predictions for multipoint OFDM downlink channels", Technical Report, Signals and Systems, Department of Engineerging Sciences, Uppsala University, May 2014, second edition March 2018. Also presented at the Swedish Communication Technologies Workshop (Swe-CTW) in Västerås, Sweden, June 2014.

II **Rikke Apelfröjd** and Mikael Sternad, "Design and measurement based evaluations of coherent JT CoMP: A study of precoding, user grouping and resource allocation using predicted CSI," *Eurasip Journal on Wireless Communications and Networking*, June 2014.

III **Rikke Apelfröjd** and Mikael Sternad, "Robust linear precoder for coordinated multipoint joint transmission under limited backhaul with imperfect CSI," *the IEEE International Symposium on Wireless Communication Systems (ISWCS)*, Aug. 2014.

IV **Rikke Apelfröjd**, Wolfgang Zirwas and Mikael Sternad, "Joint reference signal design and Kalman/Wiener channel estimation for FDD massive MIMO," Manuscript.

V **Rikke Apelfröjd**, Joachim Björsell, Mikael Sternad and Dinh Thuy Phan Huy, "Kalman smoothing for irregular pilot patterns; A case study for predictor antennas in TDD systems," Manuscript.

List of contributions not included in this thesis

VI Daniel Aronsson, Carmen Botella, Stefan Brueck, Cristina Ciochina, Valeria D'Amico, Thomas Eriksson, Richard Fritzsche, David Gesbert, Jochen Giese, Nicolas Gresset, Hardy Halbauer, Tilak Rajesh Lakshmana, Behrooz Makki, Bruno Melis, **Rikke Abildgaard Olesen**, Maria Luz Pablo, Dinh Thuy Phan Huy, Stephan Saur, Mikael Sternad, Tommy Svensson, Randa Zakhour, Wolfgang Zirwas, "Artist 4G, D1.2 Innovative advanced signal processing algorithms for interference avoidance," December 2010.

VII Mikael Sternad, Michael Grieger, **Rikke Apelfröjd**, Tommy Svensson, Daniel Aronsson and Ana Belén Martinez, "Using "predictor antennas" for long-range prediction of fast fading for moving relays," Presented at *IEEE Wireless Communications and Networking Conference (WCNC) 2012*, 4G Mobile Radio Access Networks Workshop, April 2012.

VIII **Rikke Apelfröjd**, Daniel Aronsson and Mikael Sternad, "Measurement-based evaluation of robust linear precoding for downlink CoMP," Presented at *the IEEE International Conference on Communications (ICC) 2012*, June 2012.

IX Jingya Li, Agisilaos Papadogiannis, **Rikke Apelfröjd**, Tommy Svensson and Mikael Sternad, "Performance evaluation of coordinated multi-point transmission schemes with predicted CSI," Presented at *IEEE Personal Indoor and Mobile Radio Communications (PIMRC)*, September 2012.

X Valeria D'Amico, Bruno Melis, Hardy Halbauer, Stephan Saur, Nicolas Gresset, Mourad Khanfouci, Wolfgang Zirwas, David Gesbert, Paul de Kerret, Mikael Sternad, **Rikke Apelfröjd**, Maria Luz Pablo, Richard Fritzsche, Hajer Khanfir, Slim Ben Halima, Tommy Svensson, Tilak Rajesh Lakshmana, Jingya Li, Behrooz Makki, Thomas Eriksson, "Artist 4G, D1.4 Interference avoidance techniques and system design," July 2012.

XI Tilak Rajesh Lakshmana, **Rikke Apelfröjd**, Tommy Svensson, and Mikael Sternad, "Particle swarm optimization based precoder in CoMP with measurement data," Presented at *5th Systems and Networks Optimization for Wireless (SNOW) Workshop*, April 2014.

XII Nima Jamaly, **Rikke Apelfröjd**, Ana Belén Martinez, Michael Grieger, Tommy Svensson, Mikael Sternad and Gerhard Fettweis, "Analysis and measurement of multiple antenna systems for fading channel prediction in moving relays," Presented at *the 8th European Conference on Antennas and Propagation (EuCAP)*, April 2014.

XIII Volker Jungnickel, Konstantinos Manolakis, Wolfgang Zirwas, Volker Braun, Moritz Lossow, Mikael Sternad, **Rikke Apelfröjd**, and Tommy Svensson, “The role of small cells, coordinated multi-point and massive MIMO in 5G,” *IEEE Communications Magazine*, May 2014.

XIV Annika Klockar, Mikael Sternad. Anna Brunstrõm, **Rikke Apelfröjd** and Tommy Svensson, "User-centric pre-selection and scheduling for feedback reduction in CoMP systems," *IEEE International Symposium on Wireless Communication Systems (ISWCS)*, Aug. 2014.

XV Wolfgang Zirwas, Mikael Sternad and **Rikke Apelfröjd**, "Key Solutions for a Massive MIMO FDD System," *IEEE Personal Indoor and Mobile Radio Communications (PIMRC)*, Oct. 2017.

XVI **Rikke Apelfröjd** and Mikael Sternad , "Procede d’estimation du canal entre un emitteur/recepteur et un object communicant mobile," French Patent Application no 1763263, Dec. 2017.

Contents

Abbreviations

4G 4:th Generation mobile communication system. 26, 28, 29, 67, 68, 79
5G 5:th Generation mobile communication system. 23, 28, 29, 34, 38, 42, 67

AR Autoregressive. 37, 38, 45, 46, 50–55, 57–60, 77, 101, 102, 105
ARMA Autoregressive Moving Average. 53

CDF Cumulative Distribution Function. 64, 75, 91–94
CIR Channel Impulse Response. 71
CoMP Coordinated MultiPoint. xix, xx, 23, 34, 35, 37, 39–41, 48, 65, 67, 69, 71, 75, 77, 79–86, 88, 91, 96, 105, 106
CQI Channel Quality Index. 28, 65, 88, 89
CSI Channel State Information. xix, 23, 26, 28, 34, 37, 39–42, 67, 68, 70, 73, 78, 79, 81–85, 91–95, 97, 99
CU Control Unit. 40, 79, 83, 95

DPC Dirty Paper Coding. 83

FDD Frequency Division Duplex. xiv, xx, 28, 29, 39, 42, 61, 64–71, 76, 81, 84, 97, 98, 105

GPS Global Positioning System. 97

IoT Internet-of-Things. 23
ISI Inter-Symbol Interference. 27

JB Joint Beamforming. 40, 81
JS Joint Scheduling. 40, 81, 82, 85
JT Joint Transmission. 23, 34, 35, 37, 39–41, 48, 65, 67, 69, 80–85, 88, 91, 106

LLMSE Linear Least Mean Squared Error. xxi, 39, 73–76, 105
LOS Line-Of-Sight. 24, 42, 53, 55, 98, 100, 103
LTE Long Term Evolution. 29, 34, 67, 74

MIMO Multiple-Input Multiple-Output. xiv, xx, 23, 27, 28, 32, 37–41, 44, 61, 64, 66–70, 76, 77, 81–84, 86, 89, 97, 98, 105
MISO Multiple-Input Single-Output. 84
ML Maximum Likelihood. 46, 52, 53
MMSE Minimum Mean Squared Error. xix, 33, 41

MRC Maximum Ratio Combining. 40, 68, 71–73, 77
MSE Mean Squared Error. 33–35, 41, 46, 47, 49, 84, 85, 87, 96

NLOS Non-Line-Of-Sight. 24–26, 38, 42, 55, 74, 98, 100, 103
NMSE Normalized Mean Squared Error. xix, xxi, 40, 56–60, 76, 91, 92, 98–100, 102–105

OFDM Orthogonal Frequency Division Multiplexing. 27, 28, 37, 44, 45, 56, 68, 71, 76, 86, 98, 99

RAN Radio Access Network. 81, 82

SINR Signal to Interference and Noise Ratio. 68
SISO Single-Input Single-Output. 67
SNR Signal to Noise Ratio. 28, 56–59, 71, 76, 77, 89, 100, 102, 103

TDD Time Division Duplex. xiv, xxi, 28, 29, 39, 42, 43, 48, 68, 69, 97–100, 102, 104, 107

Acknowledgements

There are a large number of people who deserve to be thanked and acknowledged in this thesis, these include, but are not limited to, those mentioned below.

First and foremost, I wish to thank supervisors Mikael Sternad and Anders Ahlén, because without them, this thesis would not have existed. Thank you for providing me with a place in the research group. A special thanks to Mikael for your support, guidance and enthusiasm.

To our secretary Ylva Johansson, thank you for making sure things run smoothly around the office.

To everyone in the Signals and Systems group, the CORE group and the FTE group, whom I have had the privilege to share my work days with. Thank you for making my days brighter. A special thank you to Simon, who has put up with sharing an office with me, to Steffi for her willingness to always discuss teaching with me and to Joachim with whom I've had the joy to work in projects during these past years.

Over the years I have had enjoyed working with people from the industry and from other universities. Some of these deserve a special mention. So to Jingya Li, Tilak Lakshamana and Nima Jamaly from Chalmers Technical University, Annika Klockar from Karlstad University, Michael Grieger, Richard Fritzsche and Fabian Diehm from TU Dresden, Konstantinos Manolakis from TU Berlin, Wolfgang Zirwas from Nokia Bell Labs and Dinh Thuy Phan Huy from Orange, thank you for enjoyable collaborations and to Tommy Svensson from Chalmers Technical University, thank you for all the advise you have provided me with.

To my family, thank you for all your love, support and encouragement. A special thank you to my parents and my in-laws who have been especially helpful babysitting these past months. To my lovely boys Kasper and Alexander I wish to thank you for filling my life with joy and to my husband Senad, thank you for always being there for me and loving me even when I'm at my worst.

Finally, to you, the reader. Whether you read the whole thing, skim through the next few chapters or only pay attention to the acknowledgements and perhaps a couple of the pictures. Thank you for your time, I hope you enjoy.

Rikke Apelfröjd

Summary of papers

When writing this thesis as a comprehensive summary, the aim has been to explain the general idea behind different concepts and to highlight those results that are of importance for further study of the subject and results that may be of importance as input to standardization of future generations mobile communication systems. Technical details such as most equations, proofs and simulation settings are found in the papers.

As a recommendation, the reader should focus on the extensive summary in Chapters 1-5 and dive into the details of the papers whenever entering a topic that is of extra interest to the reader. It is then not necessary to read the introductions of the papers, as this is mainly covered by the comprehensive summary.

In order to give an overview for those readers who may not be familiar with the area of wireless communication, Chapter 1 is kept on a very basic level reviewing some important aspects for the physical layer of wireless communication systems and the basic idea behind estimation theory. As a consequence, any reader familiar with these concepts may want to skip straight ahead to Section 2.

In terms of the included papers, denoted I-V, there is some overlap when it comes to the description of the channel models and the Kalman filter equations. The summary below is provided to guide the reader and to highlight the main points of the papers.

Comments of the author's contributions to each paper with multiple authors are stated below for each of the five included papers.

Paper I: Kalman predictions for multipoint OFDM downlink channels

This technical report provides a detailed description on how to use the Kalman filter for predicting small scale fading of channels. It extends the framework of the Ph.D. thesis [1] by Aronsson to include also channels from multiple base station sites.

Design choices, such as where to locate the filters, how to estimate the channel models and which pilot pattern to use, are discussed.

The report also includes results on the predictability of small scale fading models. It illustrates how the predictability of a channel is fundamentally limited by the fading statistics, represented by the Doppler spectrum.

The measurement based prediction results of Paper II and of [2] are highlighted and studied in detail. Some additional Normalized Mean Squared Error (NMSE) prediction statistics results that where not included in Paper II are included in this report to highlight different aspects of the prediction performance. Based on these performance results, system design issues, such as pilot patterns, intercluster interference and system delays, are discussed in the conclusion section.

The report also includes an appendix on how to generate block-fading channel models that have (instantaneous) error statistics that correspond to the one obtained in a given physical setting when Kalman predictors are applied. This method has been used in [3].

Paper II: Design and measurement based evaluations of coherent JT CoMP: A study of precoding, user grouping and resource allocation using predicted CSI

This paper investigates if Coordinated MultiPoint (CoMP) gains are realistic in real systems. The evaluations are based on measured channels, with Kalman prediction and a robust linear precoder. The linear precoder is based on a robust Minimum Mean Squared Error (MMSE) design that takes channel uncertainties into account when designing beamforming weights and uses an ad hoc method to maximizing a sumrate criterion iteratively.

The Kalman predictions provide Channel State Information (CSI) that is sufficiently accurate to achieve significant CoMP gains, even for long temporal prediction horizons (of 24 ms) at pedestrian mobility and at 2.66 GHz. For shorter prediction horizons (of 5 ms) and at 500 MHz, they would even provide good CSI at vehicular velocities.

Results show that the robustness of the proposed precoder, i.e. the fact that it takes the CSI uncertainty into account in the precoder design, provides an increase in sumrate, especially when users are randomly grouped.

Based on results of [4], which showed that user grouping is important to secure CoMP gains (compared with single cell transmission) this paper investigates different user grouping strategies. In particular, a strategy based on local scheduling, over the different resources, is suggested. It is compared both with the optimal user groups, found through a very high dimensional search of all possible combinations, and with a greedy user grouping scheme suggested in literature. The here proposed user grouping scheme performs, in terms of sum-rate, close to the optimal scheme and to the greedy scheme, at a much lower complexity.

Interestingly, the results also shows that, for a small CoMP cluster (including three single antenna base stations) when users are grouped through the suggested user grouping scheme, then the zero forcing precoder achieves similar CoMP gains as the proposed robust linear precoder.

The reader who has read Paper I can skim through Sections 2-3 and 6.3 as well as Appendix A.

The author has done the majority of the work.

Paper III: Robust linear precoder for coordinated multipoint joint transmission under limited backhaul with imperfect CSI

In this paper the robust linear precoder that is proposed in Paper II is extended to handle constraints on backhaul capacity. The aim is to ensure that the losses in CoMP gains, due to less backhaul capacity, are decreased by avoiding to design the precoder under the faulty assumptions that all channels can be used. The suggested solution is to include the backhaul constraints in the minimization criterion, via penalty terms.

Results show that if the backhaul constraints are handled as suggested, then the loss in CoMP gains is lower than if the constraints on backhaul capacity are not considered in the precoding design. The difference in loss is especially high for cell edge users, which are the users that need CoMP most and therefore have the most to loose from backhaul constraints.

The reader who has read Paper II can skim through Sections 2-3.1 and Appendix A.

The author has done the majority of the work.

Paper IV: Joint reference signal design and Kalman/Wiener channel estimation for FDD massive MIMO

In this paper, a strategy to use non orthogonal, coded, pilots in order to decrease the overhead problem that comes with deploying massive MIMO in an FDD system is proposed. It is based on a framwork suggested by Zirwas, Amin and Sternad [5].

The aim is to obtain a large reduction in the pilot overhead, as compared to orthogonal pilots, in downlinks of systems that may use both massive MIMO and CoMP. The general principle behind the proposed pilot design is based on that only a limited number of channel components are strong, as seen for a perspective from a single user. As channels from antennas located at the same base station tend to have equal strength, a design elements must be introduced to ensure this basic property. In Paper IV this is achieved by introducing a fixed grid of beams which directs the transmitted power into different directions, hence causing different channel gains at the user.

The proposed pilot code design is such that any user within the system will be able to estimate up to its K strongest channel components, where K is the number of available pilots. It has the benefit that it does not need to be re-optimized whenever a new user enters the system.

The performance in terms of channel estimation NMSE is evaluated based on Linear Least Mean Squared Error (LLMSE) filtering and Kalman filtering. A non optimal reduced Kalman filter which only estimates the relevant channel components (which are the strongest channel components at a specific users) is proposed in order to limit the computational complexity, and is evaluated.

System simulation results, in a cluster of nine base stations, show that the channel NMSE with coded non orthogonal pilots becomes around 5 dB worse than when using orthogonal pilots, at a much reduced pilot overhead. The resulting performance degradation becomes insignificant if maximum ratio transmit beamformers are designed based on channel estimates with the attained quality.

The reader who has read Paper I can skim through Sections 3.2 and appendices C.2-D.

The simulation environment, which is based on an open source environment developed by the Fraunhofer Heinrich Hertz institute [6], was created in collaboration with Wolfgang Zirwas. The author is responsible for simulating pilots measurements and implementing the different estimations algorithms, while Zirwas designed the parts relating to simulations of the radio channels and designing the fixed grid of beams.

Paper V: Kalman smoothing for irregular pilot patterns; A case study for predictor antennas in TDD systems

In this paper the Kalman filter is used to obtain smoothed interpolation estimates of the downlink channels of a TDD system, based on uplink channel estimates.

In order to perform smoothing, future measurements of the channel need to be available to the filter. This can be achieved for vehicular user equipment by placing an antenna, a predictor antenna, in front of a second antenna, the main antenna, on the roof of a vehicle in the direction of travel. The predictor antenna will then experience the channel before the main antenna and can hence collect "future" measurements of the channel for the main antenna.

Interpolation of the uplink channel needs to be performed over the duration of downlink slots, in which no uplink pilots are available. The quality of the interpolation performance influence the quality of the channel estimates of the downlink slots on which downlink transmission and beamforming is based. A good interpolation scheme will allow a longer downlink slot duration to be used for mobile users.

Evaluations based on measurements show that interpolation through Kalman smoothing of the downlink channels helps to improve the channel estimate such that the downlink slots can have durations that correspond to 0.6-0.7 of the carrier wavelength in space. If channels are only extrapolated, then this is reduced to 0.2-0.3 of the carrier wavelength.

The work of this paper was carried out in close collaboration with Joachim Björsell who has been responsible for pre processing of the measurements, while the author has been responsible for the calculations and simulations regarding the smoothing algorithm.

1. Introduction

Since the shift of the millennia wireless communication has moved from mainly supporting phone calls and occasional data transmission to and from mobile user devices to support large data rates including tracking data, cloud services and streaming services. The customers of today's wireless communication systems require constant connection and are not satisfied when data rates drop. Especially not during the commute to work.

The increasing demands require wireless transmission systems that can support a large number of data hungry users that may be densely located and/or travelling at high velocities. In addition, it is very likely that future wireless systems need to support not only the traditional devices that are directly controlled by a user such as a mobile phone or a computer, but also more or less autonomous devices that communicate amongst each other, the so called Internet-of-Things (IoT), causing the number of user equipment to increase.

In order for future systems to support both user equipments with high data demands and user equipment with low latency demands, a flexible system structure that supports a large number of transmission techniques is required. Strategies for improving spectral efficiency that are currently a part of the in standard include Multiple-Input Multiple-Output (MIMO) transmission, channel information based scheduling and adaptive modulation and coding. Candidate techniques for future standardization include Coordinated MultiPoint (CoMP) Joint Transmission (JT) and massive MIMO. All these techniques have in common that they need accurate Channel State Information (CSI) available at the transmitter side. The estimation of such CSI is crucial for high data throughput, however it is important that the quest for accurate CSI does not come at the cost of not being able to support bursty traffic.

The work presented in this thesis will focus on evaluating methods to obtain the CSI required for CoMP JT and massive MIMO, and highlighting results that can be of interest when designing 5:th Generation mobile communication system (5G) networks.

1.1 The radio channel

A traditional wireless communication system consists of one or more stationary base stations that transmit and receive data to and from a large number of user equipments. Some of the user equipments are mobile, while others are stationary. The base station (and sometimes the user equipment too) will

generally have multiple antennas. These can be used to increase the probability of success (by transmitting or receiving the same message over multiple antennas) or to direct the transmitted or received signal energy.

When a radio signal is transmitted, whether it is directed or is sent out omnidirectional, the energy of the signal will spread. A bit simplified, this can be described as the information carrying sinusoidal radio signal splitting up in multiple rays that each propagates into a different direction and interacts with the environment through reflection, refraction and by loosing energy to the medium that it travels through. An illustration of this is shown in Figure 1.1. Here, a radio signal is transmitted to a mobile phone (a piece of user equipment) on ground level from a base station situated on the roof of a tall building. Only a very small fraction of the transmitted radio frequency power will reach the intended destination. Figure 1.1 illustrates some of the paths of the radio waves that reach the mobile. The signal can be modelled as multiple rays (which in the example are reduced to four rays to make them easily distinguished) [1]. One of these rays refracts over the roof of the building and then takes a direct route to the phone, while the other rays travel in different directions and reflect one or more times on buildings before reaching the phone. As the figure illustrates, the rays that travel a longer way are weaker once they reach the mobile phone, and the one ray that is reflected from the rooftop of one of the low buildings has its strength further weakened due to the radio signals propagation through vegetation.

Any part of the radio signal that reaches the receiver in a straight path from the transmitter (without having reflected or refracted on the path) is called a Line-Of-Sight (LOS) component, whereas any part of the radio signal that has had its direction changed during the propagation to the receiver is called a Non-Line-Of-Sight (NLOS) component. Similarly, a channel with a strong LOS component and only weak NLOS components is called a LOS channel and a channel with a relatively insignificant LOS component is called a NLOS channel.

The multiple rays in Figure 1.1 will add up either constructively or destructively at the mobile user depending on the relative phase shifts of the sinusoids. As the phase shifts depend on the lengths of the paths each component has travelled, the received power will differ for different locations in space. Figure 1.2 shows an illustration of how the strength of a radio channel may look in space for a typical NLOS channel when the signal transmitted is narrowband, i.e. when it spans a small frequency interval. The standing wave pattern that appears is referred to as a small scale fading pattern. In a LOS scenario, the fading pattern will in general be much smoother. Depending on

[1] This "ray tracing" way of modelling radio wave propogation is an approximation of the exact solution, which would be obtained by solving Maxwell's equations within exactly known and specified boundary conditions.

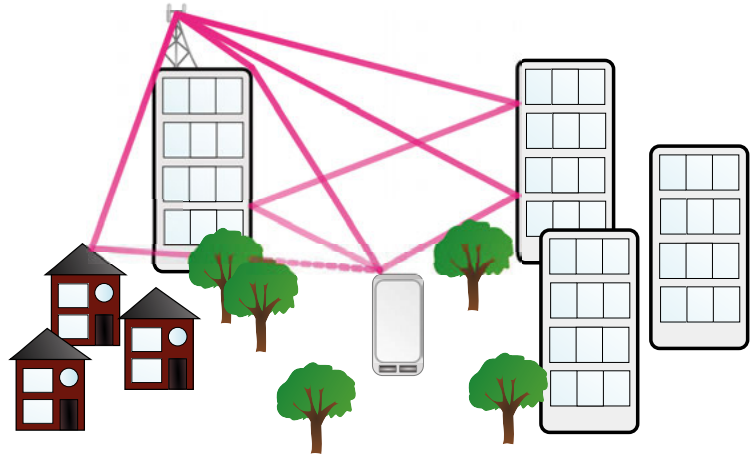

Figure 1.1. An illustration of a multipath channel. The strength of the radio signal at different points in space is indicated by the intensity of the color and the dashed line indicates a significantly weaker strength.

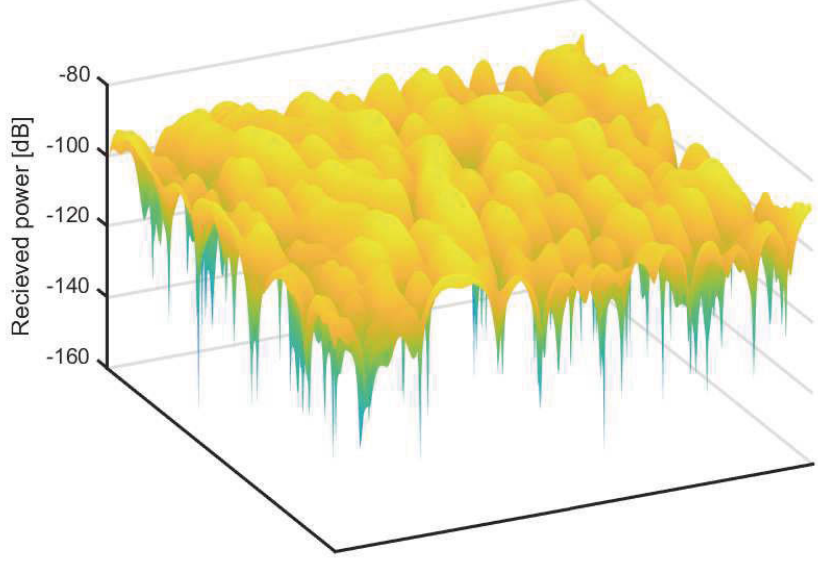

Figure 1.2. An illustration of a standing spatial wave pattern of a typical scalar urban NLOS narrowband channel. The received energy on the horizontal plane of a signal depends on the spatial location of the user equipment.

where the user equipment is situated the strength of the received power will vary.

The scale on which the fading pattern varies is directly related to the wavelength of the radio signal, which is denoted the carrier wavelength. In general the large power dips, also known as fading dips, will occur a couple of times per wavelength in an urban NLOS environment. As an example, at a carrier frequency of 2.6 GHz (which is a common frequency for the 4:th Generation mobile communication system (4G) network), the carrier wavelength is approximately 11.5 cm and fading dips will occur approximately every 4-6 cm.

1.1.1 The quest for channel state information

The description of how the transmitted signal changes as it propagates through space is referred to as the radio channel and can be represented either by an impulse response, or by a frequency response[2]. Narrowband signals are described as signals where the frequency response at a given position can be represented by one single complex number, with the absolute value representing the damping of the transmitted signal and the angle representing the phase shift. Knowledge of the channel is referred to as CSI. The CSI may include anything from a general statistical model of the set of channel properties that is consistent with a given set of background information and measurements, to a specific estimate of the channel frequency response along with information on its accuracy.

CSI is important for a number of reasons. As an example, imagine that some mobile user equipment is travelling through the standing wave pattern illustrated in Figure 1.2. Then the received signal strength at that user will vary. If a base station and a user attempt to communicate while the user is in a fading dip, then there is a large risk that the data will be lost, if it is transmitted only over a single pair of antennas, and only at this frequency. However, if the base station has knowledge of when these dips occur, then it can schedule all communication between that user equipment and the base station to when the channel gain is high. Whenever a fading dip occurs at a potential transmission frequency and time, the base station can instead choose to communicate with a different user equipment on that frequency resource and in that way increase the spectral efficiency of the system. The task of selecting which users to serve on which resources is known as scheduling.

The fading pattern and fading dips can also be modified to some extent to make transmission more favourable. For example, when a base station is equipped with several antennas, then the channels between the base stations

[2]For applications within wireless communication the physical channel can be approximated by a linear time varying dynamic system with very high accuracy, so a time varying impulse response is sufficient to describe the radio channel. Transmitter and receiver processing may add non-linearities to the total model.

antennas and an antenna at the user equipment will have different (although in general correlated) fading patterns. If the base station is aware of the phase shifts of each channel, then it can adjust the phase shifts of the radio signals from each transmit antenna to ensure that these will add up constructively at the user equipment, thus lowering the depth and the spatial density of the fading dips and improve the overall channel gain.

Similarly, when the user equipment transmits a radio signal to the base station, the receiver can weight and combine measurements from different antennas to improve the quality of the received signal.

These are only a few examples of MIMO techniques that not only can increase the data throughput, but also can allow a base station to transmit simultaneously to many users within the same frequency bands. The latter is enabled by adjusting the resulting standing wave pattern of the received signals at each user's location such that only the part of the transmitted signal intended for that particular user will be added up constructively and thus have a high receive power, while the parts of the transmitted signal that are intended for other users are left non-amplified, or are made to add up destructively.

1.1.2 Orthogonal frequency division multiplexing: A brief overview

In an Orthogonal Frequency Division Multiplexing (OFDM) system a broadband signal is created as several narrowband signals that are superpositioned into one time limited signal, denoted one OFDM symbol, before being transmitted over the radio channel. Each of these narrowband signals, often referred to as subcarriers, can then be used to encode separate pieces of information, or messages. By adjusting the frequency band of the narrowband signals based on the time duration of the OFDM symbol, modulated narrowband signals can be made orthogonal over the symbol time such that, under ideal assumptions, the messages encoded on different subcarriers will not interfere with each other. Under realistic assumptions, the system and receiver can be designed to ensure that interference between subcarriers is very small, if the transmitter and receiver are synchronized in time and frequency with sufficient accuracy. Likewise, the system is often designed to ensure that interference between subsequent OFDM symbols in time, Inter-Symbol Interference (ISI), can be considered negligible.

In an OFDM system, a single subcarrier over the duration of a single OFDM symbol is referred to as a time-frequency resource or simply *resource*. Just as the channel changes over time, it may also change over frequency. A channel can either be flat fading, with constant channel gain (although different phase) over all subcarriers, or it can be frequency selective, in which case the gain varies over different subcarriers. At the base station, a scheduling algorithm will be used to assign resources to each user equipment within the system.

Most utilized scheduling algorithms are based on some CSI, which may consist of the complex-valued channel gains or simply a Channel Quality Index (CQI). The CQI may include information of the Signal to Noise Ratio (SNR) of the subcarrier, or simply information on which subcarriers have channels that are above a given SNR threshold. The scheduler will aim to schedule messages on the resources where the channels of a user are good.

The time and frequency spread of a resource is often designed to ensure that the channel associated with it can be described by a single complex-valued scalar h. Likewise, the part of the transmitted and received signals that are associated with the time-frequency resource can also be described by complex-valued scalars, here denoted s and y respectively. The received message can then be described as the product of the transmitted signal and the channel with some additional additive noise and interference. If multiple users will be scheduled on the same resource, then multiuser MIMO techniques are used to ensure that each user equipment receives the messages intended for it. Such techniques will be further discussed in Chapter 4.

For further reading on the topic of OFDM the interested reader may refer to e.g. the works of [7], which gives a thorough theoretical background to the topic, [8], which explains the implementation of OFDM in the current 4G system, and [9] which describes the implementation of OFDM in the future 5G system.

1.1.3 Uplink and downlink

The radio channel of radio systems that have a fixed infrastructure of base stations is separated into uplink and downlink. Over the uplink, the user equipment transmits information to the base station and over the downlink the base station transmits information to the user. By using different resources for uplink and downlink, strong self-interference, i.e. that the weak received signal is interfered by its own strong transmitted signal, on the same time/frequency resources, is eliminated.

The amount of resources that are allocated to uplink and downlink respectively is a design choice that depends on how much information that is anticipated to be transmitted over each link. As surfing and streaming services become more common, it is likely that the uplink will be allocated less resources than the downlink as illustrated in Figure 1.3.

There are two common ways of separating uplink and downlink, both of which are illustrated in Figure 1.3. In FDD systems, uplink and downlink transmit simultaneously but in different frequency bands, whereas in TDD, the full bandwidth is utilized for both uplink and downlink, however the two are separated in time.

These designs both have their advantages and disadvantages. For example, as the frequency response of the channel varies in different bands, separate

Figure 1.3. Uplink and downlink resource allocation for FDD and TDD systems respectively.

channel estimations are required for uplink and downlink in FDD systems. In contrast, in TDD systems the uplink and downlink frequency responses will generally be similar for the same position in space - with some differences due to using different transmit and receive filters in the two links, which will introduce differences due to hardware imperfections in the equipment. This property is called channel reciprocity. On the other hand, low latency requirements could be easier to handle in an FDD system. If, for example, an automatic control system needs a small piece of data within a short time frame, but the system has just switched to an uplink slot, then there is a good chance that the information will be invalid by the time the system reaches its downlink slot. A potential remedy is to create flexible uplink downlink slots where users that have low latency demands may have very short switching times between uplinks and downlink. However this does place higher flexibility demands on the system and will create increased interference between different nodes.

The current 4G Long Term Evolution (LTE) systems are based on FDD with exception of the Chinese 4G LTE which is based on TDD. The paired spectral bands currently used for 4G and earlier systems will likely remain FDD spectra for a foreseeable future, whereas any new spectra that will be used for 5G systems will most likely mainly be TDD based, with a flexible uplink/downlink slot allocation, in order to adjust the resources depending on application.

1.2 Channel estimation

In order to estimate the channel required for, e.g., resource scheduling and transmission design, some resources are reserved for the base station and/or user equipment to transmit pilot signals. These are signals that are known to both the user and the base station. By measuring the received signal within a pilot resource, the channel frequency response can be estimated.

As a very simple example, consider a sinusoidal signal where information bits are coded into the amplitude and phase shift (with respect to some reference time) which are represented by the absolute value and a phase angle re-

spectively, of the complex number s, called a symbol. Furthermore, consider that this signal is transmitted through a time-invariant narrowband channel, where the complex-valued frequency response of the channel h describes how the amplitude and the phase of the transmitted signal is altered during propagation through the channel. Then the amplitude and phase of the received signal can be represented by the absolute value and the phase angel respectively of a complex-valued number y_d where

$$y_d = hs.$$

Now assume that prior to transmitting the bits represented by s, a pilot signal with the same frequency as the sinusoid carrying information bits was transmitted. We let the complex-valued p represent the known phase and amplitude of the transmitted pilot signal and $y_p = hp$ represent the amplitude and phase of the received pilot signal. As the pilot is known at both transmitter and receiver, it can be used to find the channel through the relation $h = y_p/p$, and by extension also the transmitted signal on the receiver side through

$$s = \frac{y_d}{y_p} p.$$

The example above gives the basic reasoning behind pilot-based channel estimation. However it is not an accurate representation of reality. In a more realistic scenario both the pilot measurements and the received data symbol y_d will be subjected to noise, both from the hardware equipment and from interfering signals from other transmissions e.g. from neighbouring frequency bands and/or from neighbouring transmitters. In addition, the channel may not be static. In particular, if the user equipment is mobile, then the channel that affects the pilot signal will differ from the channel that the data carrying signal experiences to some extent depending on the user mobility, the fading pattern (Figure 1.2) and the time delay between the two signals.

A more general way to approximate the pilot measurement at a single receive antenna is through

$$\mathbf{y}_\tau = \Phi_\tau \mathbf{h}_\tau + \mathbf{n}_\tau. \tag{1.1}$$

Here $\mathbf{y}_\tau$ and $\mathbf{n}_\tau$ are complex-valued column vectors of dimension K consisting of measurements and measurement noise respectively during a time window indexed by an integer τ. Each element in these represent a separate set of measurement and measurement noise, e.g. from different subcarriers. The elements of the complex-valued channel column vector $\mathbf{h}_\tau$ of dimension $N_{tx} \cdot K$ represent the channel frequency responses from the N_{tx} transmit antennas at the K time-frequency locations of the measurements. The $K \times N_{tx} \cdot K$ matrix Φ_τ is filled with pilot symbols that represent all the signals transmitted on each antenna over each of the K resources.

The problem of channel estimation is the problem of finding an estimate $\hat{\mathbf{h}}_{\tau+m}$ of the channel vector at a time window indexed by $\tau + m$ using as much

of the available information about the noise $\mathbf{n}_\tau$, the measurement $\mathbf{y}_\tau$, the pilot matrix Φ_τ and the relationship between the channel vectors $\mathbf{h}_\tau$ and $\mathbf{h}_{\tau+m}$ as realistically possible. In addition to the exact values of the pilot matrix and channel measurements, the available information often consists of first and second order statistics, i.e. mean values, covariance matrices and autocorrelation functions. Often there are also past channel measurements available.

1.2.1 Pilot design

The expression (1.1) is flexible as it allows us to choose the structure of the pilots. Three different types of pilot designs will be considered throughout this thesis.

Resource orthogonal pilots

Resource orthogonal pilots means that each transmit antenna is allocated individual time-frequency resources to transmit pilots. When a pilot resource is allocated to one antenna every other antenna must be silent. This pattern creates no inter-antenna interference.

An example with $K = 2$ and $N_{tx} = 2$ is

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} h_{11} \\ h_{12} \\ h_{21} \\ h_{22} \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \end{bmatrix}, \tag{1.2}$$

where h_{ij} is the channel at resource i from antenna j. Thus, antenna 1 sends its pilots only in resource 1 and antenna 2 sends its pilot only in resource 2. We may here directly obtain the channel estimates

$$\begin{aligned} \hat{h}_{11} &= y_1 = h_{11} + n_1, \\ \hat{h}_{22} &= y_2 = h_{22} + n_2. \end{aligned} \tag{1.3}$$

We obtain no direct measurement of h_{12} and h_{21}, but assuming that the channel from one antenna is equal to all the K transmission resources, we may use the estimates

$$\begin{aligned} \hat{h}_{12} &= \hat{h}_{11} = y_1 \\ \hat{h}_{21} &= \hat{h}_{22} = y_2. \end{aligned} \tag{1.4}$$

Code orthogonal pilots

Code orthogonal pilots allow more than one antenna to transmit pilots on the same pilot resources. However, the structure of the pilots are such that a sequence of pilots transmitted from one antenna on the given resources is orthogonal to a sequence of pilots transmitted on the resources by a another

antenna. In order to achieve this, the number of available resources must be at least equal to the number of antennas. Use of code orthogonal pilots are in general inferior to resource orthogonal pilots as it creates interference between the antennas on each resource K that cannot in general be fully cancelled at the receiver.

An example with $K = 2$ and $N_{tx} = 2$ is

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} h_{11} \\ h_{12} \\ h_{21} \\ h_{22} \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \end{bmatrix}, \tag{1.5}$$

If channels from one antenna are equal in both resources, $h_{11} = h_{12} = \bar{h}_1$ and $h_{21} = h_{22} = \bar{h}_2$, then we have a system with two unknowns and two equations, with unique solution for $n_k = 0$

$$\begin{aligned} \hat{\bar{h}}_1 &= y_1 + y_2, \\ \hat{\bar{h}}_2 &= y_1 - y_2. \end{aligned} \tag{1.6}$$

However, in general, if $h_{11} \neq h_{12}$ or $h_{21} \neq h_{22}$, we have an under determined system of equations, with no unique solution. We can then not estimate all four channels h_{ij} based on measurements at time τ only. If we in such case use the (erroneous) hypothesis $h_{11} = h_{12}$ and $h_{21} = h_{22}$ to produce the estimates (1.6), then estimation errors will be inevitable even in the noise free scenario.

Non orthogonal pilots

Non orthogonal pilots also allow multiple antennas to transmit on the same pilot resources but without the restriction that the pilot sequences must be orthogonal.

In the example above, the pilot matrix may then be

$$\Phi_\tau = \begin{bmatrix} p_1 & p_2 & 0 & 0 \\ 0 & 0 & p_3 & p_4 \end{bmatrix}, \tag{1.7}$$

where p_i are arbitrary but known complex numbers. Using non orthogonal pilots generally decreases the performance but it comes with the benefit of reducing pilot overhead, which is important in systems with a large number of antennas, e.g. massive MIMO systems, where it is desirable to use $K < N_{tx}$.

More details and comparisons between the different pilot designs are given in in Sections 3.5.1, 4.1 and 4.2.

1.2.2 Linear estimation

We saw in the examples above that estimating the channel vector $\mathbf{h}_\tau$ for multiple transmit antennas based on pilot measurements $\mathbf{y}_\tau$ at time step τ only will

in general represent an estimation problem with more unknowns than constraints. It is natural to increase the number of constraints by using additional measurements, in particular measurements that were already obtained at previous time steps.

In linear estimation the vector $\mathbf{h}_{\tau+m}$ is estimated through a weighted sum of all available measurements up to time step τ

$$\hat{\mathbf{h}}_{\tau+m|\tau} = \sum_{i=0}^{\tau} \mathbf{W}_i \mathbf{y}_i = \begin{bmatrix} \mathbf{W}_0 \dots \mathbf{W}_\tau \end{bmatrix} \begin{bmatrix} \mathbf{y}_0 \\ \vdots \\ \mathbf{y}_\tau \end{bmatrix} = \mathbf{W}\mathbf{y}. \tag{1.8}$$

Here $\tau+m|\tau$ is used to denote the estimate of the channel vector at time $\tau+m$ provided measurements up until time τ. The weighting matrices $\mathbf{W}_i$ are based on some or all of the available statistics of the channel and are chosen to minimize some criterion. An advantage to linear estimation compared to non-linear estimation is that linear estimation often requires much lower computational complexity.

The most common criterion to minimize in estimation theory is the Mean Squared Error (MSE), i.e.,

$$E[|\mathbf{h}_{\tau+m} - \hat{\mathbf{h}}_{\tau+m|\tau}|^2], \tag{1.9}$$

where $|\cdot|$ is used to represent the euclidean norm of a vector and $E[\cdot]$ denotes the expected value. It is well known that the optimal solution to the Minimum Mean Squared Error (MMSE) problem of finding a linear estimator $\mathbf{W}$ in (1.8) that minimizes the MSE (1.9) is given by the causal Wiener filter [10].

For calculating a Wiener filter, a statistical correlation model must be available for the correlation between different elements of the channel vector $\mathbf{h}_\tau$ in (1.1), and of the noise $\mathbf{n}_\tau$. By using this correlation information, a unique minimum MSE estimate is produced, also in cases as the example (1.6) with $h_{11} \neq h_{12}$ where a unique exact algebraic solution cannot be obtained.

A disadvantage of the Wiener filter is that finding the weight matrix $\mathbf{W}$ in (1.8) requires inversion of the covariance matrix of the vector $[\mathbf{y}_0^T, ..., \mathbf{y}_\tau^T]^T$. As this covariance matrix grows with every new measurement, the complexity associated with the matrix inversion will very soon become infeasible, unless we give up on using all past data, and instead use a limited sliding time window.

To lower the complexity of the estimator, a Kalman filter can instead be used. A Kalman filter is a recursive version of a Wiener filter that utilizes a state space model of the temporal correlation of the channel. The Kalman filter has the advantage that, for a wide sense stationary system, it will converge such that the more computational demanding processing can be calculated off-line, and hence the complexity associated with updating the estimate $\hat{\mathbf{h}}_{\tau+m}$ given a new measurement can be kept relatively low [1]. For this reason, the work in this theses is mainly based on the Kalman filter, which will be

described in further detail in Chapter 3, with Chapter 4 bringing up potential 5G system applications for the filter.

Whether the Kalman or the Wiener filter is used, the statistic of the channel, in the form of covariance matrices and/or autocorrelation functions must be estimated. Such estimations will always introduce some model errors, which to some extent destroy the optimality of the filter. A disadvantage to the Kalman filter compared to the Wiener filter is that model errors will be introduced in two steps. First, by estimating the covariance matrices and/or autocorrelation functions, and second when using these to estimate a state space model of the channel. The effect of these model errors on the resulting MSE is discussed further in Section 3.4.

1.2.3 Filters, predictors and smoothers

Depending on if the integer m in (1.8) is zero, positive or negative, the estimate is called a filter estimate, a prediction or a smoothed estimate respectively. The difference between these three is how much measurement data is available, at time step $\tau + m$.

The filter estimate requires measurement data up until the time of the estimate. This is useful when the channel has not changed much between the time the latest pilot measurement was received and the time the channel estimate is used. Let us consider the example presented in Section 1.1 with a carrier wavelength of 2.6 GHz, with users moving at 5 km/h (pedestrian speed) and time delays of up to 1 ms. As explained above, the dips of Figure 1.2 will then be approximately 4-6 cm apart. As a user moves through the standing wave pattern it will have travelled 1.4 mm in the time between receiving the pilot measurement and the time of using the channel estimate. Over this short distance, the channel will only have changed slightly, and the estimation errors due to this change are generally small, so the filter estimate will suffice for most pedestrian applications. However, for higher velocities or longer time delays this will not longer be true.

As pilots take up resources that would otherwise be used for data transmission it is of interest to transmit them only when required. For example, in LTE, the CSI reference signals, which are downlink pilots used for estimating channels from multiple antennas, are transmitted with an interval of at least 5 ms. Delays can also be introduced for other reasons, e.g. in a system with multiple base stations that are cooperating to transmit messages to the same users, through so called CoMP JT. Then information needs to be shared over backhaul links and this could potentially take up to tens of ms. As the system delays increase, the channel will change to a greater extent and the CSI provided by the filter estimate will be *outdated*. A similar effect is obtained at higher user velocities. In such scenarios predictions are required. Section 4.2

focuses on channel prediction for scenarios with long delays and slowly moving users when predictions are used for CoMP JT.

A way of describing the small scale fading of a single narrowband radio channel h_τ, from the perspective from a user that is moving through a standing wave pattern as the one in Figure 1.2, is by the channel's autocorrelation function $R(t) = E[h_\tau h^*_{\tau-t}]$ or by a Doppler spectrum. The Doppler spectrum is given by the Fourier transform of the autocorrelation function. The width and shape of the Doppler spectrum affects the predictability of a radio channel. This will be discussed further in Section 3.4.

In order to obtain a smoothed estimate, measurement data from both past and future, relative to the time of interest, are required. An example when this may be available is if the receiver is in no rush to detect its signals and can wait for the next pilot measurement before using all available pilot measurements to estimate the channel and hence detect the transmitted symbol. A second application for channel smoothing is described in Section 4.3 and includes the use of a predictor antenna, a concept which will be described briefly in the next subsection and in more detail in Section 4.3.

The challenge with long range channel prediction and how to solve it

It is intuitive that access to more measurement data also should provide better estimates. Hence, the smoothed estimate outperforms the filtered estimate in terms of MSE and the filtered estimate in turns outperforms the prediction. It is likewise intuitive that the quality of the prediction decreases as the prediction horizon m in (1.8) increases. It has been shown in e.g. [1], that a prediction horizon beyond a few tenths of the carrier wavelength in space is infeasible using linear predictors. The reasons behind this will be explored more in Section 3.4.

For a user equipment that moves through a standing wave pattern generated by a stationary transmit antenna and fixed reflecting or scattering objects, a required prediction horizon of L seconds is equivalent to a prediction horizon in space in terms of carrier wavelengths

$$\frac{Lv}{\lambda} = \frac{Lvf_c}{c} \text{ [wavelengths].} \tag{1.10}$$

where v is the velocity in m/s, λ is the carrier wavelength, c is the speed of light and f_c is the carrier frequency. As an example, at $f_c = 3.5$ GHz predicting 10 ms ahead in time would at a velocity of 30 m/s correspond to prediction over a distance of 3.5 wavelengths in space.

A way to push the prediction horizon beyond that of a few tenths of the carrier wavelength for vehicle users is by utilizing a predictor antenna. The concept is illustrated in Figure 1.4. Here, two antennas are positioned on the roof of a bus, aligned along the direction of travel. The forward antenna, which is denoted the predictor antenna, transmits or receives pilots (depending on the system). From these pilots, the channel is estimated and this filter estimate

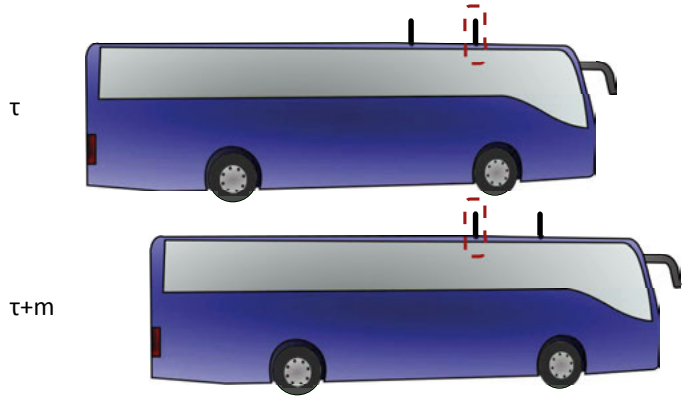

Figure 1.4. Illustration of the predictor antenna concept. In this example, a bus is equiped with two antennas aligned along the direction of travel. At a time τ the forward antenna, denoted predictor antenna, either transmits or receives a pilot (depending on the system) allowing for a filter estimate of the channel at the location marked by the red square. At the time $\tau + m$ the rearward antenna, denoted main antenna, has entered the same location, so the filter estimate based on the prediction antenna at time τ can be used as a channel prediction for the main antenna.

can then be used to design a transmitter that transmits to the rearward antenna, denoted main antenna, at a later time when it has reached the same position in space as where the predictor antenna was at the time of pilot transmission.

The predictor antenna concept can be used to gain access to future measurements of the channel, relative to the position of interest, if the antennas are sufficiently separated. This also allows for a smoothing estimate, based on the pilots received or transmitted by the predictor antenna, to be used as channel predictions for the rearward antenna.

2. Contributions

For many wireless transmission schemes, accurate CSI at the transmitter of a downlink is crucial to achieve desirable gains. Such schemes include adaptive modulation and coding [11], channel aware scheduling [12] and multiuser MIMO transmission, e.g. zero forcing, [13].

For this reason, channel estimation has been a topic of interest for wireless communication for a long time. Particularly, the use of linear filtering is useful, as it helps to keep computational complexity at bay.

2.1 The Kalman filter

Kalman estimators and predictors for OFDM MIMO channels, based on Autoregressive (AR) models for fading statistics, were proposed by Aronsson in [1]. In Paper I and Paper II these methods are extended to multi-antenna and multi-site downlink channels and are evaluated for use in CoMP JT. An overview of how to obtain the AR model and the noise covariance matrices that specify a Kalman filter is given in Paper I, along with discussions of design choices of both the modelling and the Kalman filter.

2.1.1 Factors limiting the predictability of the radio channels

The Kalman filter can be used not only as a tool for estimation, but also as a theoretical tool for exploring the attainable estimation accuracy under various assumptions. One such long-standing problem are the basic reasons for the very limited predictability of fading radio channels that are measured under realistic conditions. Results based on channel sounding measurements in [14] and in [1] have consistently found that it is in general hard to predict the small scale fading of a component (either in the time domain or in the frequency domain) more than a couple of tenths of wavelengths ahead in space.

On the other hand, the physical time varying channel generated by moving through a stationary fading environment will be strictly band limited: The support of the Doppler spectrum is constrained to $\pm$ the maximum Doppler frequency $f_d = v/\lambda$, where v is the velocity and λ is the carrier wavelength. It is known [15, 16] that strictly band limited signals can be infinitely well predictable from past noise free measurements.

If this is the case, then why is it practically impossible to predict fading channels over multiple wavelengths? Is it because a predictor is forced to

extrapolate based on noisy past channel estimates, or is the fundamental reason something else?

The question of what factors limit the predictability is discussed in Section 3.4 and in Paper I, using properties of AR modelling and Kalman based prediction based on AR models.

It turns out that the noise level on past estimates has a relatively small influence on the error of long range channel predictions. The fundamental reasons for lack of predictability is instead that finite order *models* of the fading process, that are based either on a finite set of training data or on noisy training data, cannot be band limited. This lack of band limitation in the model very efficiently destroys long-range predictability.

Models based on time limited data sets will essentially see the physical fading channel through a time window, and this window smears the Doppler spectrum. Models based on noisy training data will have a noise floor in their Doppler spectrum. In both cases, the models will lack infinite peaks (pure sinusoids) which would theoretically be perfectly predictable.

A starting point for the present investigation are interesting results obtained by Baddour and Beaulieu [17] on one step prediction errors of AR models of fading processes obtained from noisy training data sets.

Here, the method of [17] is used to approximate a band limited theoretical Doppler spectra for NLOS channels with uniformly distributed scatterers in two dimensions [18] by a high order AR model that is obtained by adding a small regularization term to the autocorrelation of the channel at zero time shift. Based on this, various factors that affects the predictability of the radio channel are evaluated.

Furthermore, results in Section 3.4 abd Paper I show that the measurement noise associated with the pilot measurements is of less importance than other factors when it comes to the range of predictability of a channel. The first order effect is the quality of the training data, i.e. how much broadening and smearing it induces into the Doppler spectrum of the estimated fading model. The second order effect is the model order. When the autocorrelation function of the channel is perfectly known, with except for some white noise that slightly alters the term to the autocorrelation of the channel at zero time shift in the same way as the regularization term described above, then a higher model order leads to more accurate predictions. However, a too high model order will result in worse channel estimate whenever the channel statistics changes over time. Results here supports what was found in [1, 19]; that linear prediction beyond a couple of tenths of the carrier wavelength has very poor accuracy.

2.2 Kalman filters for potential 5G applications

In Chapter 4, linear estimation, and in particular Kalman filtering is used to investigate the potential of different MIMO transmission techniques for 5G

systems in cases where channel estimation is challenging. In particular, solutions are proposed and investigated for the following three problems: Massive MIMO for FDD systems, coherent CoMP JT with potentially long backhaul delays and massive MIMO for TDD systems with long downlink slots and vehicular users.

2.2.1 Estimation of massive MIMO FDD channels

A scheme for downlink channel estimation for massive MIMO in FDD systems must solve two main problems. First, we have a potentially very large set of channel components that need to be estimated without introducing an unreasonably large overhead. Second, the solution must support users at a large number of potential locations, with very different conditions in terms of channel gains and fading. The scheme that is presented in Paper IV and summarised in Section 4.1 is one for pilot design and channel estimation for an FDD implementation. The goal is that the overhead is made to scale with the number of channels that will be relevant for a terminal, which is typically in the range 5-30, in systems with hundreds of antenna elements.

The primary key property that is used is that when the channel components have varying average gain, then each user only has to estimate the strongest channel components, as seen from that user. If different users will have different strongest channel components then estimating only their strongest channel components will lead to an insignificant decrease in the multiusers scheduling gain. Signals from antennas located at different sites will in general have large differences in received power. For antennas located at the same site, the average channel gains should, on the other hand, be similar. Therefore, it is crucial to introduce some system design elements to reduce this similarity between channels for collocated antennas.

The proposed framework has four components:

1. Antennas will be structured into a fixed *grid of beams* where each beam is wideband and controlled by an effective or virtual antenna port. At any given user position, only fractions of the antenna ports will then have strong signal, so only a fraction of their channels need to be estimated.
2. Downlink pilots will be transmitted as non orthogonal pilots, using *coded pilot sequences*. The codes are designed such that they provide unique pilot patterns for each of a potentially very large number of antenna ports within a cooperation area. The size of the pilot blocks (the coded sequence length) is selected proportional to the number of channel components that need to be estimated for a typical user.
3. *Correlation* over time, space and frequency is utilized by a linear Linear Least Mean Squared Error (LLMSE) estimator or by a Kalman filter to improve the CSI quality.

4. Use of *cycling sequences of pilot codes* ensures good estimations regardless of the user position, by weighted time averaging of estimation errors that may be caused by the non orthogonality of some of the pilot codes, as seen from one user.

System simulations using the Matlab based, open source, Quadriga channel simulator, developed by the Fraunhofer Heinrich Hertz institute [20, 6] show that by using the fixed grid of beams and coded non orthogonal pilots, channel estimations with a reduced Kalman filter provided an average channel estimate Normalized Mean Squared Error (NMSE) of 5 dB worse than what could be achieved with resource orthogonal pilots. The resource orthogonal pilots would, however, cause a much higher overhead. The beamforming gains through Maximum Ratio Combining (MRC) were not noticeably affected by this NMSE reduction.

2.2.2 Channel prediction to overcome backhaul delays in coordinated multipoint systems

There are many challenges and hurdles to overcome before downlink CoMP can be fully deployed, especially for coherent JT. The work presented in Papers II-III and summarised in Section 4.2 focuses on three of these challenges.

First, data sharing over backhaul links will cause time delays between the channel estimation and the payload transmission. CoMP decisions will then be based on outdated CSI. How this affects the potential CoMP gains depends on the scheme used, i.e. Joint Scheduling (JS), Joint Beamforming (JB) or JT, and on the network architecture, i.e. centralized, distributed or semi distributed, see e.g. [21, 22]. It turns out that the CSI quality is especially important for coherent JT in a centralized architecture, which is also the solution that would provide the largest potential gains. It is therefore of great importance to find methods to improve the CSI, under long backhaul delay constraints, in order to make coherent CoMP JT feasible.

Data sharing over backhaul links places high demands on the backhaul capacity, especially for JT. In realistic systems, these demands may not always be met. Then the information available at the Control Unit (CU), which is a logical entity that makes the CoMP decisions, will be limited. Furthermore, the possibility to share payload data between the base stations may also be limited. If the CoMP design does not handle these *backhaul constraints*, then potential CoMP gains may be lost.

Second, a nontrivial problem for coherent CoMP is that of which users to serve jointly on each specific transmission resource. If these are selected carefully then there is a potential for large multiuser diversity gains, as was shown for multiuser MIMO in e.g. [23, 24]. However the complexity of a search through all possible user groups and all resource allocation possibilities grows combinatorially with the number of users to choose from. It is desirable to

have a simple scheme that is low in complexity and that preserves the multiuser diversity gains of the optimal solution.

Third, the messages that are to be transmitted to the users need to be precoded to mitigate inter user interference within the cooperation cluster. In a system with perfect CSI this can be achieved by channel matrix inversion, which provides the zero forcing precoder developed for MIMO, see e.g. [25]. However, there is a risk in using zero forcing for CoMP JT, even in the presence of perfect CSI as it is only designed to minimize intracluster interference, but not to weight this against intercluster interference. An option is then to use an MSE criterion which includes these terms as well as the intracluster interference [26].

MSE criteria are attractive as they generally have analytical solutions. However, in practice it is often more useful to maximize over a weighted sumrate criterion, as this is closer to the desired end performance. Such an optimization poses a multidimensional nonconvex optimisation problem.

The work presented in Papers II-III and summarised in Section 4.2 aims to adress these challenges through three main contributions

1. The problem of the outdated CSI is partly counteracted by channel predictions through Kalman filtering.
2. A very low complexity method for user grouping and resource allocation is proposed and evaluated. This method provides very close to optimal CoMP groups and resource allocation, in the case of the sumrate criterion. It places no extra requirements on backhaul capacity.
3. A low complexity robust linear precoder design is proposed. This is a robust MMSE design that offers a flexible tool to optimize over a few parameters with respect to an arbitrary criterion, e.g. the sumrate. The robust design takes channel uncertainties due to prediction errors and quantization into account. It can also include backhaul constraints in a flexible way.

For the evaluations, pilot sounding channel measurements collected by Ericsson in Stockholm have been used. The results show that channel predictions can ensure CoMP JT performance gains for system delays of tens of miliseconds at pedestrian speed, or, if the system delays are shorter, then at even higher speeds.

2.2.3 Channel smoothing for TDD systems with predictor antennas

For pedestrian users, the required prediction horizon is usually within the 0.1-0.3 carrier wavelength limitation, whereas for high mobility users, a longer prediction horizon is often required.

This limit of prediction horizon can be circumvented by the predictor antenna concept, introduced in subsection 1.2.3. It uses an extra antenna, *pre-*

dictor antenna, placed in front of the *main antenna*, e.g. on the roof of the vehicle, to scout the channel that will later be encountered by the main antenna, as originally proposed in [27], see Figure 1.4.

In an FDD system, the predictor antenna would receive known pilots in the downlink and based on these estimate the channel. If pilots are transmitted often and the two antennas are sufficiently spaced, then one of these estimates can be used as a prediction of the channel for the main antenna. However, at high user velocities, the frequency with which pilots must be transmitted for the CSI not to be outdated may cause undesirable overhead.

For a TDD system, the predictor antenna would transmit known pilots in the uplink subframes, that are used for channel estimation on the network side. Assuming channel reciprocity, these channel estimates are then used to calculate predictions of the channel to the main antenna during a subsequent downlink subframe. The uplink/downlink ratio of the TDD frame might be adjusted so that the downlink transmission of the main antenna occurs close to a position where the predictor antenna already has measured the channel, as proposed in [28]. However, such a scheme would require individually adaptable uplink/downlink ratios based on the velocity of each user which is problematic from a system perspective.

Instead of transmitting a large number of pilots in the FDD case and relaying on individually adaptable uplink/downlink ratios in TDD systems, interpolation can be used for any given uplink/downlink ratio to generate channel estimates for the gaps in the pilot sequence.

The results in Paper V, which are summarised in Section 4.3, use the Kalman smoother to interpolate the channel estimates of the predictor antenna. It focuses on the TDD scenario as it will be more challenging since the pilots are irregularly spaced in time and the longest durations without any pilots (due to downlink frames) are longer than for the FDD case. This motivates us to study the performance of interpolation schemes in this setting.

The Kalman smoother used is based on a two-filter approach, using two state space models. The measurement based simulation results provided show that in the presence of predictor antennas, when CSI is interpolated based on Kalman smoothing, the CSI quality within the downlink slot can be improved significantly. As a result the downlink slot duration can be extended significantly compared to when only extrapolation of CSI into the downlink slot is utilized. The results can be used to determine how the flexibility of developing 5G standards in terms of reference signal rates and TDD subframe duration, can be used for vehicular users, some of which may use predictor antennas.

The concept has been validated based on a small subset of a larger set of channel sounding measurements collected by TU Dresden in Dresden, Germany. The measurements used were selected to represent three distinct fading scenarios: LOS, NLOS with a Jakes like Doppler spectrum and NLOS with a flat Doppler spectrum. Results show that with Kalman smoothing, the frequency with which pilots for channel estimation need to be transmitted, at the

beginning and at the end of the gap due to a downlink slot, can be pushed to what corresponds to a spatial resolution of 0.6-0.8 of the carrier wavelength. This relaxes constraints, so that a relatively long TDD downlink duration can be used in transmissions to mobile vehicular users that utilize predictor antennas.

3. The Kalman filter

The Kalman filter was developed by Rudolf Kalman [29]. It takes a Bayesian approach to estimation, utilizing a priori information of the channel in the form of a state space model of the channel's dynamics and a prior estimate of the state vector. Whenever a measurement of the channel is available then an a posteriori channel estimate can be computed. In this chapter, the basic principles of the Kalman filter, along with some design choices and various sources of errors due to noise and mismodelling are addressed. For details on these topics, the reader is referred to Paper I, and in the case of pilot designs, also Paper IV.

3.1 Background and related work

The Kalman filter has been suggested along with the Wiener filter as a good candidate for channel estimation [1, 30, 31, 32, 33, 34, 35]. Thorough tutorials on the topic of estimation based on linear filtering in general and Kalman filtering specifically can be found in e.g. [10, 36, 37].

The Kalman filter is essentially a recursive Wiener filter which uses a state space model and a state vector (as a prior) to account for all prior knowledge of the channel statistic and previous states. These are combined with new measurement information through a Baysian estimation [10, 36]. The Kalman filter has an advantage to the Wiener filter that, due to the use of the state space model, the Kalman filter can account for past channel measurements at lower complexity than that of the Wiener filter. It can also handle known time varying properties of linear dynamic models and signal statistics.

In [30] a one dimensional Wiener filter was used to estimate channels in the frequency domain. The 2D Wiener filter was mentioned as a candidate to estimate OFDM channels over both time and frequency by [31], while [1] focuses on using the Kalman filter for estimation of fading and frequency selective OFDM channels. The latter provides a tutorial on how to estimate state space models for OFDM MIMO channels and investigates the potential of these for channel prediction.

In this thesis Kalman filtering, prediction and smoothing of OFDM channels are carried out in the frequency domain. An alternative is to filter in the time domain, which is also described in [1]. Time domain prediction is investigated in e.g. [38]. In [39], the Kalman predictions presented here was

compared to the time domain predictions for real measured channels. The results showed no big difference between the two alternatives.

Channel predictions through linear filtering has been investigated in e.g. [1, 19]. In these works, the predictability of radio channels seems to be in the range of a couple of tenths of the carrier wavelength. However, theoretical results suggest that a band limited signal should be infinitely predictable [15, 16]. There are two large differences between the theoretical results and the practical results. First, perfect prediction requires perfect knowledge of the channel statistics, however in reality this must always be estimated. Noise will be included in the training data, i.e. the data that is used to estimate the model, and even though the influence of noise can be suppressed by averaging, some error will always remain. In addition, the training data will always be time limited. Second, although the physical fading radio channel may be band limited (to the maximum Doppler frequency), the effective radio channel as seen from the user is not. As the received signal is always time limited, it cannot be band limited in the Doppler frequency domain.

In [17] it was shown that, for the theoretical channels described by [18] and [40], the one step prediction error could be made very small if only Gaussian noise is included. Section 3.4 will discuss what this means for a longer prediction horizon, and we will see that even when the channel is almost band limited (only a small part of the spectral energy is outside the band limit) long range predictability is infeasible.

3.2 Mathematical background

The majority of the work included in this thesis focuses of estimating fading OFDM radio channels through Kalman filtering, Kalman prediction and Kalman smoothing. The Kalman filter requires a state space model of the signal that is to be estimated, which in this case consists of narrow band channels (subcarriers) represented by complex numbers. The state space model represents the assumed statistics of the small scale fading of each subcarriers. For this section we shall assume that these models are known, and then in Sections 3.3 and 3.4 will go into the issues with estimating models.

Let us assume that a vector of N complex-valued zero mean narrow band channels $\mathbf{h}_\tau \in \mathbb{C}^{N\times 1}$ can be described by a discrete time wide sense stationary AR state space model

$$\begin{aligned} \mathbf{x}_{\tau+1} &= \mathbf{A}\mathbf{x}_\tau + \mathbf{B}\mathbf{w}_\tau, \\ \mathbf{h}_\tau &= \mathbf{C}\mathbf{x}_\tau, \\ \mathbf{Q} &= E[\mathbf{w}_\tau \mathbf{w}_\tau^*], \\ \mathbf{\Pi} &= E[\mathbf{x}_\tau \mathbf{x}_\tau^*], \end{aligned} \tag{3.1}$$

which describes its evolution over time. Here, $\mathbf{w}_\tau \in \mathbb{C}^{N\times 1}$ is a zero mean vector of white Gaussian noise with time independent autocorrelation $\mathbf{Q}$, denoted process noise, and $\mathbf{x}_\tau \in \mathbb{C}^{Nn_{AR}\times 1}$ is the state vector which is assumed to be independent of $\mathbf{w}_\tau$. The matrices $\mathbf{A} \in \mathbb{C}^{Nn_{AR}\times Nn_{AR}}$, $\mathbf{B} \in \mathbb{C}^{Nn_{AR}\times N}$ and $\mathbf{C} \in \mathbb{C}^{N\times Nn_{AR}}$ are time independent state space matrices and n_{AR} is the model order of an AR model that is used to represent the temporal correlation of a single channel component, which will be further discussed in Section 3.3. The integer time index τ represents the OFDM-symbols when pilots may be sent.

In a situation where $\mathbf{x}_\tau$ is perfectly known, the extrapolation of the state vector one step into the future, $\mathbf{x}_{\tau+1}$, would consist of two terms: The term $\mathbf{A}\mathbf{x}_\tau$, which would be known, and the term $\mathbf{B}\mathbf{w}_\tau$, for which only the second order moments, represented by $\mathbf{Q}$, would be known. This last term represents what is new and unknown at time $\tau+1$ in the state vector $\mathbf{x}_{\tau+1}$.

As the process noise is assumed to be zero mean Gaussian, the Maximum Likelihood (ML) estimate of the second term would be an all zero vector, and the one step prediction error of $\mathbf{h}_{\tau+1}$ would become $\mathbf{C}\mathbf{B}\mathbf{w}_\tau$. As was shown in [15, 16], the one step prediction error for band limited signals can be forced to be zero. The next state vector must then be perfectly described by a number of past state vector and so $\mathbf{Q} = 0$ in such a state space model.

Furthermore, let us assume that a $\mathbf{y}_\tau \in \mathbb{C}^{K\times 1}$ represents a vector of pilot measurements at time τ at K transmission resources where only known pilots were transmitted. The entries of $\mathbf{y}_\tau$ may represent measurements at different pilot bearing time-frequency resources and/or at different receive antennas. The measurement process is modelled by

$$\mathbf{y}_\tau = \Phi_\tau \mathbf{h}_\tau + \mathbf{n}_\tau \tag{3.2}$$

where $\Phi_\tau \in \mathbb{C}^{K\times N}$ is a pilot matrix and $\mathbf{n}_\tau \in \mathbb{C}^{K\times 1}$ represents the sum of noise and interference, e.g. from base stations not considered in the cluster, which is assumed zero mean with covariance matrix $\mathbf{R} = E[\mathbf{n}_\tau \mathbf{n}_\tau^*]$ of full rank. It will, unless otherwise specified, be denoted measurement noise.

It is worth noting that the estimations that are evaluated here are solely based on pilot measurements of the channel. These estimates could be further improved upon by using the transmitted data symbols as well. In such a case, the pilot based channel estimate will first be used to determine which symbol has been sent, e.g. through ML detection. As only a discrete number of symbols can be sent (the number is determined by the modulation format), the ML estimated symbol can then be used as a pilot to re-estimate the channel. This process can be repeated iteratively if desired.

The aim is now to produce an MSE optimal estimate $\hat{\mathbf{x}}_\tau$ of the state vector $\mathbf{x}_\tau$ (and thereby of $\mathbf{h}_\tau = \mathbf{C}\mathbf{x}_\tau$), based on the measurement $\mathbf{y}_\tau$ and on all other available relevant information.

The state space model (3.1) represents the a priori information, along with an estimate of the state space vector $\hat{\mathbf{x}}_{\tau-1}$ and an error covariance matrix of

this estimate

$$\mathbf{P}_{\tau-1} = E[\tilde{\mathbf{x}}_{\tau-1}\tilde{\mathbf{x}}_{\tau-1}^*], \tag{3.3}$$

where the estimation error $\tilde{\mathbf{x}}_{\tau-1} = \mathbf{x}_{\tau-1} - \hat{\mathbf{x}}_{\tau-1}$ is uncorrelated to the estimate. As the state space vector is a sum of weighted independent complex-valued white Gaussian zero mean vectors, through (3.1), it is in itself a zero mean complex-valued white Gaussian vector. Therefore, in the case that no other estimate is available, the best a priori guess of the state space vector would be its mean value, i.e. an all zero vector and $\mathbf{P}_{\tau-1} = \Pi$. In fact, this is a way to initiate the filter. Other options for initiating the Kalman filter are discussed in Paper I.

Based on the a priori information and the measurement (3.2), an a posteriori MSE estimate is given by the recursive set of matrix difference equations, know as the Kalman equations

$$\hat{\mathbf{x}}_\tau = \mathbf{A}\hat{\mathbf{x}}_{\tau-1} + \mathbf{K}_\tau(\mathbf{y}_\tau - \mathbf{J}_\tau\mathbf{A}\hat{\mathbf{x}}_{\tau-1}), \tag{3.4}$$

$$\mathbf{P}_\tau = (\mathbf{I} - \mathbf{K}_\tau\mathbf{J}_\tau)(\mathbf{A}\mathbf{P}_{\tau-1}\mathbf{A}^* + \mathbf{B}\mathbf{Q}\mathbf{B}^*), \tag{3.5}$$

$$\mathbf{K}_\tau = (\mathbf{A}\mathbf{P}_{\tau-1}\mathbf{A}^* + \mathbf{B}\mathbf{Q}\mathbf{B}^*)\mathbf{J}_\tau^*(\mathbf{R} + \mathbf{J}_\tau(\mathbf{A}\mathbf{P}_{\tau-1}\mathbf{A}^* + \mathbf{B}\mathbf{Q}\mathbf{B}^*)\mathbf{J}_\tau^*)^{-1}, \tag{3.6}$$

$$\hat{\mathbf{h}}_\tau = \mathbf{C}\hat{\mathbf{x}}_\tau, \tag{3.7}$$

where $\mathbf{J}_\tau = \Phi_\tau\mathbf{C}$ and * denotes the conjugate transpose. The matrix $\mathbf{K}_\tau \in Nn_{AR} \times K$ is known as the Kalman gain and through (3.4) it adjusts the a prior estimate, which is given in form of the one step prediction based on the previous state vector estimate $\mathbf{A}\hat{\mathbf{x}}_{\tau-1}$, by weighting the error of the one step prediction of the measurement, that is produced by using $\hat{\mathbf{h}}_\tau = \mathbf{C}\mathbf{A}\hat{\mathbf{x}}_{\tau-1}$ which is given by $\mathbf{y}_\tau - \Phi_\tau\hat{\mathbf{h}}_\tau = \mathbf{y}_\tau - \mathbf{J}_\tau\mathbf{A}\hat{\mathbf{x}}_{\tau-1}$.

The focus here will be on the information that we gain from equations (3.4)-(3.7) while the details on how to derive (3.4)-(3.7) are given in e.g. [1, 10, 36, 37]. For more information on Bayesian inference, the reader is referred to [41].

The Kalman filter is optimal in the sense that it weights all available information in order to get the best estimate (with respect to the MSE). Therefore, the conclusions that can be drawn from equations (3.4)-(3.7) are in accordance with what most people intuitively would expect.

If the previous estimate of the state vector is very inaccurate (compared to the the power of the measurement noise and the process noise), then $\mathbf{P}_{\tau-1}$ will be large and, by (3.4) and (3.6), the estimate of the state vector will be based almost purely on the measurement. If the opposite is true and the state vector estimate is accurate, then the Kalman gain will depend on the ratio between the powers of the measurement noise $\mathbf{n}_\tau$ and the process noise $\mathbf{w}_\tau$.

When the process noise is large compared to the measurement noise, then the a posteriori estimate will be based on the measurement to a larger extent and if the measurement noise is large compared to the process noise, the a posteriori estimate will be primarily based on the a priori information. Both a

large measurement noise vector and a large process noise vector will increase the error of the a posterior estimate, through (3.5)-(3.6). Likewise, a large error in the prior estimate will in general cause a large error in the posterior estimate.

3.2.1 Predictions and smoothing

In the pilot examples in Section 1.2.1 it was shown that some hypothesis is needed for how channel components that are not directly related to time-frequency resources with pilots are related to the measurements $\mathbf{y}_\tau$. In a Kalman filter, the dynamic state space model (3.1) fulfills that purpose. Wherever no direct measurements are available, model based extrapolation or interpolation is used.

When there is no available pilot measurement at time step τ, e.g. if the channel that is to be estimated is in the future or if no uplink pilots are transmitted during e.g. a downlink frame in a TDD system, this can be described by setting the pilot matrix Φ_τ to an all zero matrix. Then equation (3.2) gives $\mathbf{y}_\tau = \mathbf{n}_\tau$ and since $\mathbf{J}_\tau = \Phi_\tau \mathbf{C} = 0$, the Kalman gain will by (3.6) be an all zero matrix. Then the estimated state vector obtained only from an extrapolation of $\hat{\mathbf{x}}_{\tau-1}$ and the corresponding error covariance matrix become

$$\begin{aligned} \hat{\mathbf{x}}_{\tau|\tau-1} &= \mathbf{A}\hat{\mathbf{x}}_{\tau-1}, \\ \mathbf{P}_{\tau|\tau-1} &= \mathbf{A}\mathbf{P}_{\tau-1}\mathbf{A}^* + \mathbf{B}\mathbf{Q}\mathbf{B}^*. \end{aligned} \tag{3.8}$$

Here, the notation $\tau_1|\tau_2$ is introduced to denote an estimate of the state vector at time τ_1 given measurements up until time τ_2.

The one step prediction in (3.8) is simple to extend to a multistep prediction by continuing to assume that the pilot matrix is an all zero matrix. We then obtain

$$\begin{aligned} \hat{\mathbf{x}}_{\tau|\tau-m} &= \mathbf{A}^m \hat{\mathbf{x}}_{\tau-m}, \\ \mathbf{P}_{\tau|\tau-m} &= \mathbf{A}^m \mathbf{P}_{\tau-m} \left(\mathbf{A}^*\right)^m + \sum_{i=0}^{m-1} \mathbf{A}^i \mathbf{B}\mathbf{Q}\mathbf{B}^* \left(\mathbf{A}^*\right)^i. \end{aligned} \tag{3.9}$$

The corresponding channel prediction estimate and prediction error covariance matrix of the channel estimate are given by

$$\hat{\mathbf{h}}_{\tau|\tau-m} = \mathbf{C}\hat{\mathbf{x}}_{\tau|\tau-m}, \tag{3.10}$$

and

$$\Gamma_{\tau|\tau-m} = E[\left(\mathbf{h}_\tau - \hat{\mathbf{h}}_{\tau|\tau-m}\right)\left(\mathbf{h}_\tau - \hat{\mathbf{h}}_{\tau|\tau-m}\right)^*] = \mathbf{C}\mathbf{P}_{\tau|\tau-m}\mathbf{C}^*, \tag{3.11}$$

respectively.

A great advantage to the Kalman filter, as compared to other linear filters is that the error covariance matrix of the state vector is part of the "package deal",

i.e. it is calculated as part of the Kalman equations (3.4)-(3.7). This information can be used for example in a multi-antenna transmit precoding stage to ensure that poor channel estimates have less impact on the final solution than accurate channel estimates. This will be discussed further and utilized for robust precoding for CoMP JT in Section 4.2.

Prediction is required when you need to estimate the channel before you have access to measurements of it. It will in general result in a prediction that is worse than the filter estimate (3.7). If one on the other hand has the opportunity to wait with estimating the channel until some more pilot measurements are available, then a smoothing estimate $\hat{\mathbf{h}}_{\tau|\tau+m}$ for $m > 0$ can be obtained.

One use of smoothing is to improve upon the accuracy of a filter estimate. Another use, that will be the focus of Paper V, is to obtain channel estimates at time steps τ at which no pilots are available, by using both past and future measurements.

There are two standard ways to perform channel smoothing. One option is by extending the state vector and including all future states up until the point when there are no more measurements available.

A second option is to use two filters. The explicit details on how to do this are given in Paper V. The basic idea is that the first filter calculates the filter or prediction estimate of the channel based on all available measurements up until the time τ of the estimate by (3.4)-(3.7). Then a second filter uses a time reversed state model and performs a backward recursion to estimate the state vector of the time reversed system at time $\tau+1$ based on all future available pilot measurements. This is then extrapolated through the time inverted state space model into a one step backwards prediction, similar to the forward prediction (3.8). Through this, we obtain a channel estimate based on the future measurements up until a given time T which we denote $\hat{\mathbf{h}}_{\tau|\tau+1,\ldots,\tau+T}$. Assuming that this estimate has a covariance matrix $\Gamma_{\tau|\tau+1,\ldots,\tau+T}$, the combined smoothed channel estimate is given by an MSE optimal weighting of the backward and forward estimates

$$\hat{\mathbf{h}}_{\tau|\tau+T} = \Gamma_{\tau|\tau+T}(\Gamma_{\tau|\tau}^{-1}\hat{\mathbf{h}}_{\tau} + \Gamma_{\tau|\tau+1,\ldots,\tau+T}^{-1}\hat{\mathbf{h}}_{\tau|\tau+1,\ldots,\tau+T}), \tag{3.12}$$

with

$$\Gamma_{\tau|\tau+T} = \left(\Gamma_{\tau|\tau}^{-1} + \Gamma_{\tau|\tau+1,\ldots,\tau+T}^{-1} - \Pi^{-1}\right)^{-1}. \tag{3.13}$$

(The term Π^{-1}, which represents prior information of the channel is included in both $\Gamma_{\tau|\tau}^{-1}$ and $\Gamma_{\tau|\tau+1,\ldots,\tau+T}^{-1}$, and hence one of these must be removed in (3.13).) For details, please see Sections 3.4.3 and 10.4 of [10].

3.2.2 Comments on complexity and the use of stationary filters

The complexity of the Kalman filter has been investigated in [1], where Table 4.1 specifies the complexity of the different operation steps in the Kalman

filter, assuming that the state model (3.1) is set up on diagonal form. From this work, we see that the highest complexity is related to calculating the Kalman gain (3.6) and the error covariance matrix (3.5).

Fortunately, an important property of the Kalman filter is that for any stable time invariant system (3.1), with time invariant Φ_τ in the measurement equation (3.2) the Kalman filter converges such that $\mathbf{P}_{\tau-1} \to \mathbf{P}_\tau$, when $\tau \to \infty$. Moreover, it will converge regardless of whether the initial error covariance matrix is known or not, i.e. even if $\mathbf{P}_0$ is not an accurate representation of the second order statistics of the estimation error of $\hat{\mathbf{h}}_0$ the filter will converge, although at a slower rate [10]. Therefore a large part of the on line complexity can be reduced by calculating both the error covariance filter and through that also the Kalman gain off-line. The error covariance matrix of the stationary filter $\mathbf{P}_f$ can be found by setting $\mathbf{P}_f = \mathbf{P}_{\tau-1} = \mathbf{P}_\tau$ in (3.5), which gives a discrete time algebraic Riccati equation

$$\begin{aligned} \mathbf{P}_f &= (\mathbf{I} - \mathbf{K}_f \mathbf{J}_\tau)(\mathbf{A}\mathbf{P}_f\mathbf{A}^* + \mathbf{B}\mathbf{Q}\mathbf{B}^*), \\ \mathbf{K}_f &= (\mathbf{A}\mathbf{P}_f\mathbf{A}^* + \mathbf{B}\mathbf{Q}\mathbf{B}^*)\mathbf{J}_\tau^* \left(\mathbf{R} + \mathbf{J}_\tau(\mathbf{A}\mathbf{P}_f\mathbf{A}^* + \mathbf{B}\mathbf{Q}\mathbf{B}^*)\mathbf{J}_\tau^*\right)^{-1}. \end{aligned} \tag{3.14}$$

Figures 3.1 and 3.2 illustrate the convergence of Kalman filters for two different types of AR models. These depict the difference between the covariance matrix of the one step prediction error when the filter error covariance matrix is calculated through the filter recursions (3.4)-(3.6) and when it is found by the stationary solution to the Riccati equation, where $\mathbf{P}_p$ denotes the error covariance matrix of the one step prediction of the stationary filter given by

$$\mathbf{P}_p = \mathbf{A}\mathbf{P}_f\mathbf{A}^* + \mathbf{B}\mathbf{Q}\mathbf{B}^*. \tag{3.15}$$

We can see that the filter converges fairly quickly, especially with a Doppler spectrum that is relatively flat. As the AR model has to be based on training data, that same training data set can be used to ensure that the filter converges off-line before it is to be used.

3.3 Model estimation

The model (3.1) must be estimated before the Kalman filter can be used. The work here is based on AR modeling based on the Yule Walker equations, see e.g. [42], as this was shown to give the best estimation performance in [1].

In general the accuracy of the model will improve the more information that is available. If the channel is approximately wide sense stationary then the more past channel data that is included in the training data, the better the AR model will be. For stationary (non-moving) users, this assumption is in general valid over a long time window. Channels to stationary or very slow moving users are also fairly easy to estimate (and predict) with very good accuracy

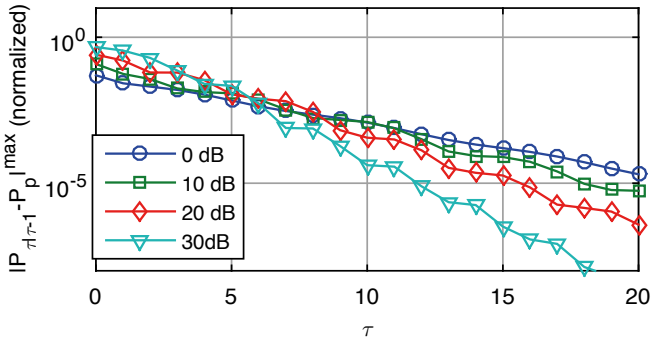

Figure 3.1. The maximum norm of $\mathbf{P}_{\tau|\tau-1} - \mathbf{P}_p$ normalized by the max norm of $\mathbf{P}_p$, where $\mathbf{P}_p$ is the one step prediction error that obeys the algebraic Riccati equation. Results are shown for different SNR of the pilot measurements by (3.2). The system in (3.1) is a fourth order AR model for a single channel tap $\mathbf{h}_\tau$ with the poles in $0.82 \pm 0.29i$ and $0.70 \pm 0.10i$, which gives an almost flat Doppler spectrum similar to that used in [40] (but not band limited). Further details are given in Section 2.3 of Paper I.

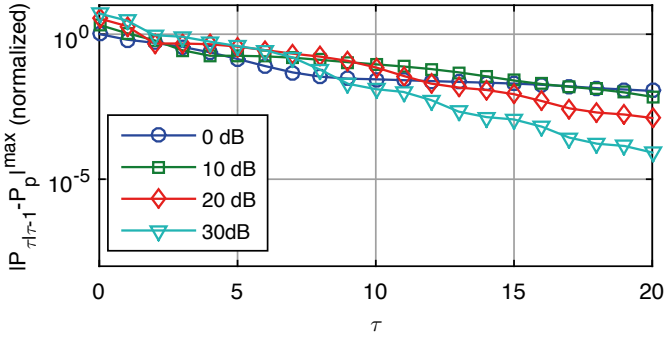

Figure 3.2. The maximum norm of $\mathbf{P}_{\tau|\tau-1} - \mathbf{P}_p$ normalized by the max norm of $\mathbf{P}_p$. Results are shown for different SNR of the pilot measurements by (3.2). The system in (3.1) is a fourth order AR model for a single channel tap $\mathbf{h}_\tau$ with the poles in $0.91 \pm 0.35i$ and $0.86 \pm 0.33i$, which gives a Doppler spectra similar to the Jakes spectrum [18] but not band limited. Further details are given in Section 2.3 of Paper I.

as their channels remain approximately constant, so many past measurements can be combined to determine their present state.

The need for more advanced estimators arises primarily when users are mobile. For these users, the small scale fading can only be considered a wide sense stationary process while the shadow fading statistics (the numbers and powers of the contributing multipaths) remains relatively constant. Through the measurements used in Papers I and II it could be observed that for pedestrian users and at a carrier frequency of 2.66 GHz, over a time of approximately on second, the assumption of a wide sense stationary system is relatively sound.

This section provides a brief overview of how to estimate the parameters of the state space model in (3.1). For the full details on how to estimate this model, please see Section 3 of Paper I.

3.3.1 Estimation of the parameters of the state space matrices of one channel element

The AR model that represents the small scale fading is estimated based on training data. What is considered training data may vary. When a new user enters a system, then the channels can be estimated by e.g. a ML estimate. These can then be used as training data to find the autocorrelation function of each individual component in the channel vector, $h_{i,\tau}$

$$R_h(t) = E[h_{i,\tau} h^*_{i,\tau-t}]. \tag{3.16}$$

Here, $h_{i,\tau}$ is an individual element of the channel vector $\mathbf{h}_\tau$.

Alternatively, all available pilot and data measurements can be demodulated and used to estimate the channels. The sequence of estimated channels can then be further smoothed and used as training data to find an AR model that in turn is used to filter the original data.

The relation between an AR model of one channel tap $h_{i,\tau}$ of order n_{AR}

$$h_{i,\tau} = -a_1 h_{i,\tau-1} - ... - a_{n_{AR}} h_{i,\tau-n_{AR}} + v_\tau, \tag{3.17}$$

and the autocorrelation function (3.16) is given by the Yule-Walker equations

$$-\mathbf{Ya} = \mathbf{z}, \tag{3.18}$$

where $\mathbf{a}$ is a vector of the AR coefficients $\mathbf{a}^T = [a_1, ..., a_{n_{AR}}]$ and

$$\begin{aligned} \mathbf{Y} &= \begin{bmatrix} R_h(0) & R_h^*(1) & \dots & R_h^*(n_{AR}-1) \\ R_h(1) & R_h(0) & \dots & R_h^*(n_{AR}-2) \\ \vdots & \vdots & \ddots & \dots \\ R_h(n_{AR}-1) & R_h(n_{AR}-2) & \dots & R_h(0) \end{bmatrix} \\ \mathbf{z}^T &= \begin{bmatrix} R_h(1) & R_h(2) & \vdots & R_h(n_{AR}) \end{bmatrix}. \end{aligned} \tag{3.19}$$

By inverting the matrix on the left hand side, the problem of estimating **a** can be solved, so that the variance of the one step prediction error, which equals the driving noise term v_τ in (3.17), is minimized.

The poles of the AR process (3.17), i.e. the roots of the polynomial

$$z^{n_{AR}} + a_1 z^{n_{AR}-1} + ... + a_{n_{AR}},$$

represent peaks in the Doppler spectrum of the estimated model. A pole close to the unit circle represents a very narrow-band peak, which could for example be a strong LOS component, whereas a pole further inside the unit circle will provide a wider peak. For a strictly band limited signal, all poles will be placed on the unit circle.

The small scale fading could also be modelled by an Autoregressive Moving Average (ARMA) process. Then the zeros of such a model could be used to suppress some parts of the Doppler spectrum. However, while the AR parameters can be found by a closed form solution of the linear problem (3.18), estimating the parameters of an ARMA model based on training data of the past channel realizations pose a non linear estimation problem, which may be solved by an iterative optimization algorithm. ML estimation is the most commonly used framework [42]. Comparisons with ARMA models are outside the scope of this thesis.

3.3.2 Modelling the correlation between channel components

When the AR coefficients are found, each individual channel element can be set up on state space form, as described in Section 3 of Paper I. By extenting the state vector to include the state vectors for all channel elements, the model (3.1) is obtained if the same AR order n_{AR} is used to describe all channel components.

The correlations between the different channel components are addressed by the covariance matrix **Q** of the process noise. Estimation of the process noise covariance matrix is a nontrivial problem, for which several solutions are proposed and compared in this thesis.

When the state space model is on diagonal form and the channel model is a perfect representation of the temporal correlation and, furthermore, the covariance matrix of the channel vector $\mathbf{R_h} = E[\mathbf{h}_\tau \mathbf{h}_\tau^*]$ is perfectly known, then the process noise covariance matrix can be found through [1]

$$\mathbf{Q} = \mathbf{R_h} \oslash \mathbf{C}(\mathbf{B1B}^* \oslash (\mathbf{1} - \mathrm{diag}(\mathbf{A})\mathrm{diag}(\mathbf{A})^*))\mathbf{C}^*, \tag{3.20}$$

where $\mathrm{diag}(\mathbf{A})$ is a column vector containing the diagonal elements of **A**, $\oslash$ denotes elementwise division, and **1** is a matrix of ones.

The expression (3.20) provides a good choice also when using imperfectly estimated state space matrices in (3.1) under special circumstances. These include the case when all channel components can be assumed to be identically

distributed, e.g. MIMO channels as in [1]. They also include cases when the channel components are uncorrelated, e.g. for different site antennas as in Paper I. The details of why these special circumstances provide a positive definite matrix for $\mathbf{Q}$ are provided in Appendix C of Paper I.

However, the channels are in general neither identically distributed nor uncorrelated, e.g. in the case of a fixed grid of beams, which will be further discussed in Section 4.1 and Paper IV. In addition, we cannot expect our estimates of the state space matrices to perfectly fit the data. Under such general conditions, the solution to (3.20) may provide an estimate of the process noise covariance matrix that is non-positive definite. Such an error will destroy the convergence of a Kalman filter.

In Paper IV, an investigation of different heuristic ways to estimate the covariance matrix of the process noise are evaluated and compared to when it is optimized through a non convex optimisation scheme of high complexity. It turns out that the optimisation tends to end up at local minima and that none of these locally optimal solutions performed better than the heuristic solutions suggested in Paper IV.

3.4 Limitations of predictability

As mentioned previously, although the physical radio channel may be band limited by the maximum Doppler frequency, the model of the radio channel is based on time limited and noisy training data. Because of this, infinite predictability is not possible. This section focuses on the predictability of a single channel component, i.e. the channel vector will only include a single element, which will below be denoted h_τ. The small scale fading of the channel is modelled by an estimated AR model.

For band limited signals the solution to the Yule-Walker equations (3.18) can be numerically difficult to find, as the matrix on the left hand side of the equation may be poorly conditioned, leading to numerical errors even for a rather low model order . One straightforward way of eliminating the numerical errors arising from the use of inverting the matrix in (3.18), which was used in [17], is by regularization. By adding a small positive number to the diagonal elements of the matrix $\mathbf{Y}$, i.e. by replacing $R(0)$ by $R(0)+\varepsilon$ where ε is a small positive number we obtain a more well conditioned matrix. Then, equation (3.18) thus becomes

$$-(\varepsilon\mathbf{I}+\mathbf{Y})\mathbf{a}_\varepsilon = \mathbf{z}, \tag{3.21}$$

where $\mathbf{a}_\varepsilon$ represents the regularized solution. This corresponds to the situation when the past channel realizations are perfectly described by the autocorrelation, except for some small white measurement noise. In real systems, the regularization term will always be present to some extent, because the effective channel that we see at the receiver side will always include some residual thermic noise, interference and other unmodelled components.

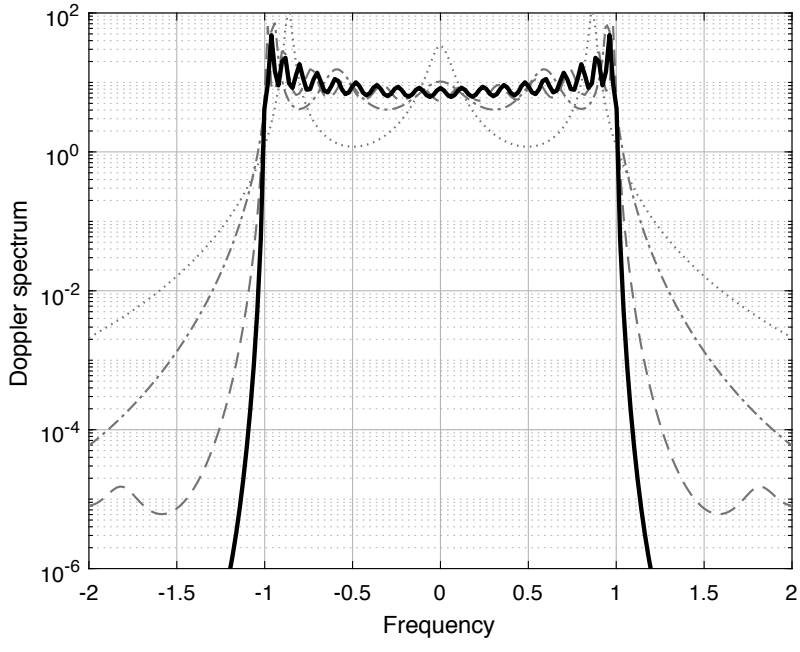

Figure 3.3. An illustration of the Doppler spectra of the $n_{AR} = 128$ order channel models that has been used to generate the simulated channels (black solid lines). This is given by solving (3.21) using $\varepsilon = 10^{-7}$ with the Jake's channel model (3.22). The figure also illustrate the Doppler spectra of channel models calculated by (3.21) using $\varepsilon = 10^{-5}$ and AR models of orders $n_{AR} = 4$ (dotted line), $n_{AR} = 16$ (dashed-dotted line) and $n_{AR} = 64$ (dashed line).

To study the effect of different factors on the feasible channel prediction horizon, a channel has been simulated using an AR model of order $n_{AR} = 128$. The model of the simulated data is given by solving (3.21) with $\varepsilon = 10^{-7}$ and where the autocorrelation function is given by the Jake's channel model described in [18] where the autocorrelation function is given by

$$R_{h,J}(t) = J_0(2\pi f_d T_p t). \tag{3.22}$$

Here, J_0 is the zeroth order Bessel function on first kind, f_d is the maximum Doppler frequency and T_p is the pilot sampling interval.

This corresponds to a NLOS scenario where the mobile user is surrounded by uniform scatters. When the channel is LOS or when it has one or a few strong and very distinct components, then the poles of the discrete time model will tend to be placed very close to the unit circle, and the resulting Doppler spectrum will be very narrow. For such scenarios, the prediction horizon can be pushed further, see [1].

Note that this means that the simulated channel is not strictly band limited, but it is close enough to approximate it as band limited in comparison to the

channel models that will be used in the Kalman filters as comparisons. The Doppler spectrum of the simulated channel is shown in Figure 3.3. This figure also shows the Doppler spectra of models that are gained from solving (3.21) for lower order channel models, $n_{AR} = \{4, 16, 64\}$ with $\varepsilon = 10^{-5}$. These three models all have Doppler spectra that are much less limited compared with the Doppler spectrum of the simulated channel. As $\varepsilon = 10^{-5}$ corresponds to a training data NMSE of -50 dB, it is highly unlikely that it would be possible to provide a better estimated model than these. In fact, training data noise corresponding to $\varepsilon = 10^{-5}$ is only reasonable for very good pilot SNR and strong correlation between channels at different OFDM subcarriers such that the noise can be even further suppressed by utilizing measurements from adjacent subcarriers.

3.4.1 The effects of noisy training data

Figure 3.4 shows how the prediction NMSE is affected by having noisy training data. Here, a channel of 1000 samples, spaced by 0.05 of the carrier wavelength (a total of 50 carrier wavelengths in space) is simulated. Based on these samples, pilot measurement signals were generated through (3.2) with a pilot SNR of 100 dB (practically noise free). The channels were then predicted through a Kalman filter using (3.4)-(3.10), where the model is provided by a model estimated through a noisy training data sequence of infinite length, i.e. through (3.21), where the autocorrelation is perfectly known and where the noise level of the training data is modelled by $\varepsilon = \{10^{-5}, 10^{-3}, 10^{-1}\}$.

The results are shown for model orders of $n_{AR} = \{4, 16, 64, 128\}$. For a realistic setting, a model order of $n_{AR} = 64$ is too high. There are two main reasons for this. The first is that the on-line complexity of the Kalman filter grows with the square of the model order, and if many channel elements are to be estimated jointly, then the complexity with $n_{AR} = 64$ will be too large.

The second reason is that the model used to estimate a given set of channel data is based on past channel data. If the models are updated frequently, then the most dominant features of the Doppler spectrum will be similar, however on a more detailed level, there will be differences. If a very high model order is selected, then the statistics of the training data will be modelled. As these details differ from the channel that is to be estimated, a high model order can actually lower the prediction performance. An investigation in [1] showed that for slowly moving users, a fourth to sixth model order was a good choice. A similar study in [43] suggested that for vehicular users at a speed of 50 km/h, a second order model order was better. Section 4.2 of Paper I shows some examples of this model order selection tradoff.

There are two main insights to gain from Figure 3.4. First, even if the autocorrelation function could be perfectly estimated except for some low noise term (with $\varepsilon = 10^{-5}$), prediction beyond half of the carrier wavelength is in-

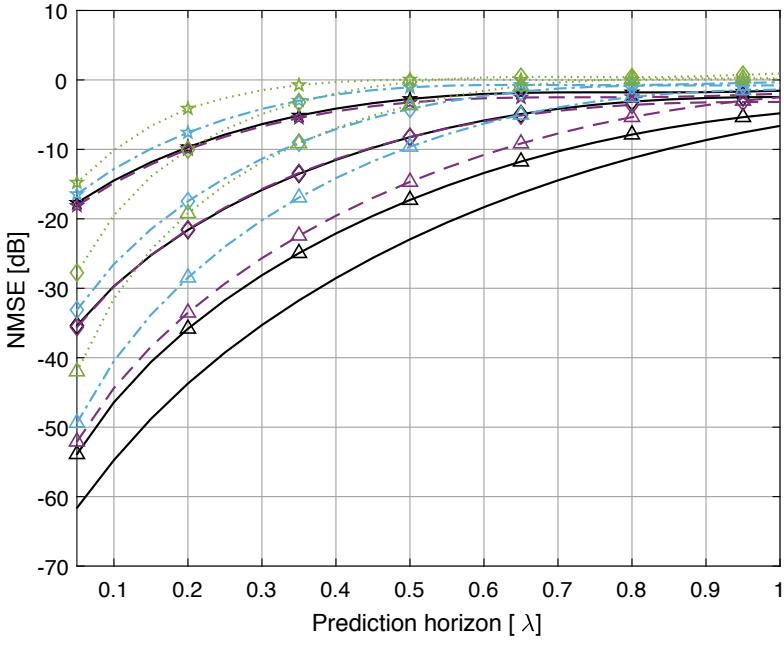

Figure 3.4. Prediction error NMSE for different model orders and training data noise levels. Here, the channel is simulated based on an $n_{AR} = 128$ order AR process found by solving (3.21) using $\varepsilon = 10^{-7}$ and the autocorrelation function of the Jake's channel model (3.22). The models used in the Kalman filters are found by solving (3.21) using $\varepsilon = 10^{-5}$ (triangles), $\varepsilon = 10^{-3}$ (diamonds) and $\varepsilon = 10^{-1}$ (pentagrams), with the autocorrelation function of the Jake's channel model (3.22). Results are shown for model orders of $n_{AR} = 128$ (solid lines), $n_{AR} = 64$ (dashed lines), $n_{AR} = 16$ (dashed-dotted lines) and $n_{AR} = 4$ (dotted lines) and with a pilot SNR of 100 dB (practically noise free). The predictability given when the original process is used as a model in the Kalman filter is added as an unmarked solid line for comparison. The prediction horizon is given in terms of the carrier wavelength λ.

feasible at reasonable model orders in a Jake's fading like scenario. Second, the gain of using a higher model order decreases as the training data becomes noisier.

3.4.2 The effects of model order and measurement noise

From (3.1) it is clear that if $\mathbf{x}_\tau$ is known, then the uncertain part of the one step prediction of the channel is $\mathbf{CBw}_\tau$. We can also see from (3.9) that if the system model is stable then the long range prediction error will depend on the process noise, but not on the pilot measurement noise, as for a stable discrete time system $\mathbf{A}^m \to 0$ as $m \to \infty$, and the impact of $\mathbf{P}_{\tau-m}$ on $\mathbf{P}_{\tau|\tau-m}$ will become negligible as m increases.

It is relevant to ask for which prediction horizons the measurement noise is influential.

To see how the SNR of the pilots affects the prediction NMSE compared to the model order, the simulations are repeated for pilot SNRs of $\{10, 20, 30\}$ dB, with $\varepsilon = 10^{-5}$. The results of these are shown in Figure 3.5.

A first observation from Figure 3.5 is that, while the pilot measurement SNR is important for the short range predictions, the model order, and hence the limitation of the Doppler spectrum is of higher importance for the long range predictions. This confirms the intuition behind equation (3.9).

3.4.3 The effect of the amount of available training data

As seen in Figure 3.3 limiting the model order smears the estimated Doppler spectrum. A similar effect is caused when the autocorrelation function is estimated based on a limited amount of training data (which is always the case for a realistic scenario). The time limitation of the training data corresponds to multiplying an infinitely long training sequence with a window function in the time domain, which will cause a convolution of the Doppler spectrum with a sinc-function, hence smearing the spectrum. We can therefore guess that the effect of having access to limited training data will cause similar effects as having a limited model order.

In Figure 3.6 the effect that having limited training data has on the predictability of a radio channel is studied. For this purpose, an additional T channel samples were simulated based on the $n_{AR} = 128$ order AR process given by solving (3.21) with $\varepsilon = 10^{-7}$ and the autocorrelation function of the Jake's channel model (3.22). From these, the autocorrelation function was estimated, and the models used for the Kalman filters were then calculated through (3.21) with $\varepsilon = 10^{-7}$, i.e. assuming a training data NMSE of -50 dB.

As the estimation of the autocorrelation varies between the different sets of training data, especially when the training data window is short, the simu-

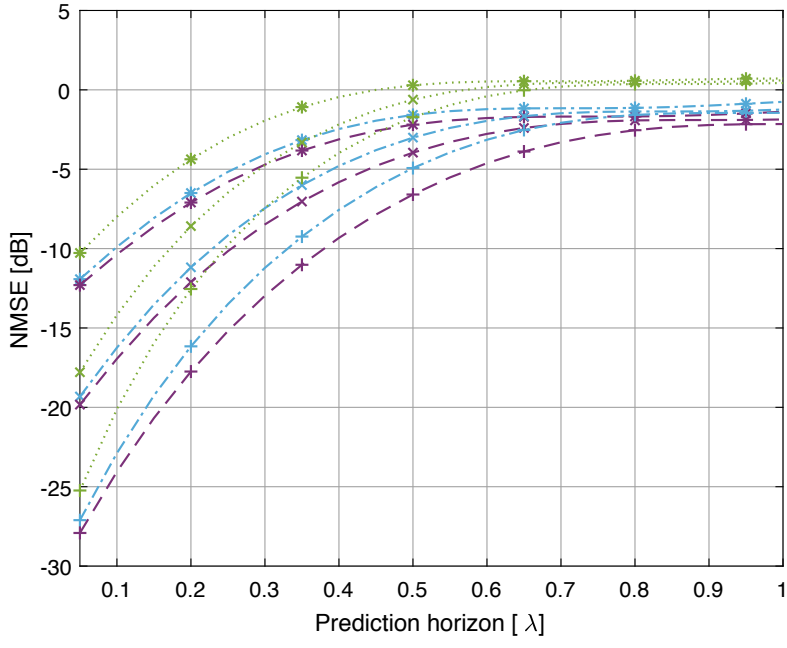

Figure 3.5. Prediction error NMSE for different model orders and pilot measurement SNR levels. Here, the channel is simulated based on an $n_{AR} = 128$ order AR process found by solving (3.21) using $\varepsilon = 10^{-7}$ and the autocorrelation function of the Jake's channel model (3.22). The models used in the Kalman filters are found by solving (3.21) using $\varepsilon = 10^{-5}$, with the autocorrelation function of the Jake's channel model (3.22). Results are shown for model orders of $n_{AR} = 64$ (dashed lines), $n_{AR} = 16$ (dashed-dotted lines) and $n_{AR} = 4$ (dotted lines) and for pilot SNR of 10 dB (stars), 20 dB (crosses) and 30 dB (pluses). The prediction horizon is given in terms of ther carrier wavelength λ.

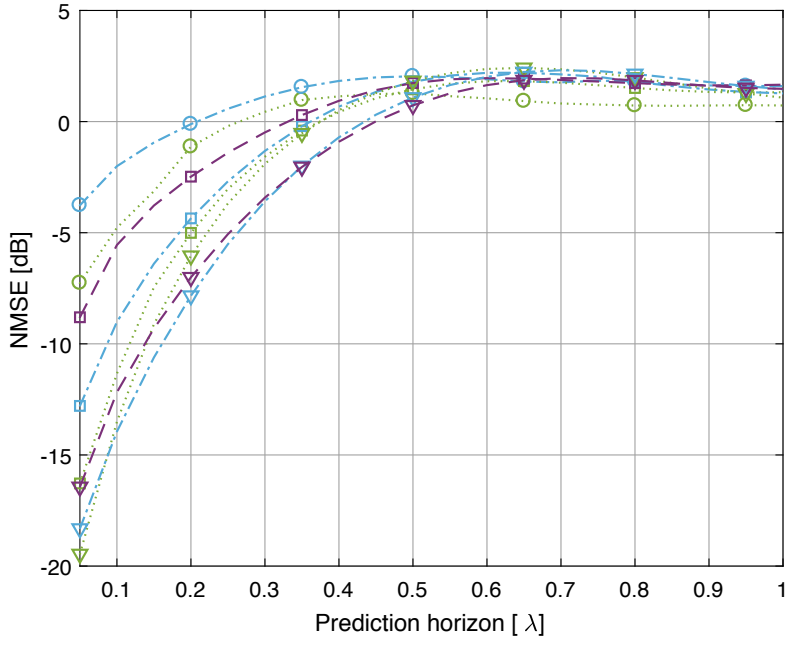

Figure 3.6. Prediction error NMSE for different model orders and pilot measurement SNR levels. Here, the channel is simulated based on an $n_{AR} = 128$ order AR process found by solving (3.21) using $\varepsilon = 10^{-7}$ and the autocorrelation function of the Jake's channel model (3.22). The models used in the Kalman filters are found by first estimating the autocorrelation function R_h in (3.22) based on a limited amount of noise free training data and then solving (3.21) using $\varepsilon = 10^{-5}$, with the estimated autocorrelation function. Results are shown for model orders of $n_{AR} = 64$ (dashed lines), $n_{AR} = 16$ (dashed-dotted lines) and $n_{AR} = 4$ (dotted lines) and for training data sets of $T = 50$ samples, corresponding to 2.5 wavelengths in space, (circles), $T = 500$ samples, corresponding to 25 wavelengths in space, (squares) and $T = 5000$ samples, corresponding to 250 wavelengths in space (triangles). The prediction horizon is given in terms of ther carrier wavelength λ.

lations were repeated 150 times and the NMSE provided in Figure 3.6 is the average over the 150 simulations.

It is clear that a limited amount of training data severely affects the predictability of the radio channel. Predictions beyond 0.2 of the carrier wavelength will not possible for a prediction NMSE below -8 dB when the training data is limited to $T = 5000$.

An interesting observation is that it is no longer necessarily the highest model order that provides the lowest prediction NMSE. For example, for a very limited set of training data (of $T = 50$ samples), it is better to choose a low order model ($n_{AR} = 4$), than to chose a higher order ($n_{AR} = 16$) [1] . The reasoning behind this is that the higher order terms of the estimated autocorrelation function will be less correct if few training data samples are available. These inaccuracies in the higher order terms of the estimated autocorrelation function will cause large errors in a high order model. It is only for a very long training data set ($T = 5000$) that it is actually beneficial to use a model order of $n_{AR} = 16$ rather than $n_{AR} = 4$.

For many vehicular applications, it is likely that the set of useful training data is very limited, as the radio channel, in these cases, can only be considered wide sense stationary over a short time window. For example, consider a vehicle moving at 50 km/h, a pilot sampling time of 5 ms and a environment that can be considered quasi static in intervals of 10 meters (which would correspond to 87 times the carrier wavelength in a system with carrier frequency $f_c = 2.6$ GHz), then we have access to 72 pilot samples to use as training data. On the other hand if the user moves at pedestrian speed, e.g. 5 km/h, we have access to a tenfold increase of training data (500 samples).

To summarize the findings of Figure 3.4-3.6, the first order effect that limits the predictability of linear channel predictions is the quality and available amount of reliable training data. This is a fundamental limitation to the predictability of a radio channel of a vehicular users, because the channel can only be assumed wide sense stationary over a limited area (which depends on the fading environment) and the training data will only be useful as long as the model based on this training data is used to predict channels within this area.

A the second order effect that limits the predictability of a radio channel is the model order. However, here it is important to know that a larger model order does not necessary mean better prediction performance. In fact, if the available amount of training data is low, then the a low model order is to prefer to a high order model.

The measurement noise of the pilot measurements comes in as a third order effect for longer prediction horizon, but is important for short prediction horizons.

[1] Note that there is no model of order $n_{AR} = 64$ for $T = 50$, as the time lag of the estimated autocorrealtion function is limited by the amount of available training data.

3.5 Design choices

This section will include a brief summary of some of the design aspects that are addressed in Section 4 of Paper I.

3.5.1 Pilot signal design and the use of coded non orthogonal pilots

Similar to e.g. [44], results of Paper I suggest that resource orthogonal pilots are to be preferred over code orthogonal pilots. However for the massive MIMO FDD downlink scenario considered in Paper IV, both of these alternatives give rise to a large overhead, if the number of pilot resources used within an approximately flat fading time frequency region equals the number of downlink channels. The non orthogonal pilot scheme that is used in Paper IV instead permits that pilots over K pilot bearing resources are linearly independent for up to K selected channels out of all the channels included in $\mathbf{h}_\tau$.

As an example, consider a downlink channel vector $\mathbf{h}_{\tau,\text{flat}}$ with $N = 9$ elements representing flat fading channels from nine antenna ports, to a single antenna receiver. Each antenna port either controls one antenna or one beam in a fixed grid of beams, at the transmitter side. Each channel is here assumed constant (flat fading) over a total of $K = 6$ pilot bearing subcarriers. As the channels are assumed flat fading, the measurement equation (3.2) is adjusted to

$$\mathbf{y}_\tau = \Phi_{\tau,\text{flat}}\mathbf{h}_{\tau,\text{flat}} + \mathbf{n}_\tau. \tag{3.23}$$

where the pilot matrix $\Phi_{\tau,\text{flat}}$ is of dimension $K \times N$. Let this pilot matrix be given by

$$\Phi_{\tau,\text{flat}} = \begin{bmatrix} -1 & 0 & 1 & 1 & -1 & 0 & -1 & -1 & 1 \\ 0 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & -1 \\ -1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 \\ -1 & 0 & -1 & 0 & -1 & 0 & 1 & -1 & -1 \\ 0 & -1 & 0 & 0 & -1 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & -1 & -1 & -1 \end{bmatrix}. \tag{3.24}$$

Furthermore, assume that a number of ≤ 6 channels, have a larger gain than the remaining channels, and are collected into a vector $\mathbf{h}_{\tau,\text{flat,rel}}$ of relevant channel components. The remaining weak (non relevant) channel components are collected in the vector $\mathbf{h}_{\tau,\text{flat,r}\bar{\text{e}}\text{l}}$. The use of the notations relevant and non relevant implies that the weak channels are not relevant for the application, and therefore do not need to be estimated.

The measurement (3.23) can then be rewritten as

$$\mathbf{y}_\tau = \Phi_{\tau,\text{flat,rel}}\mathbf{h}_{\tau,\text{flat,rel}} + \Phi_{\tau,\text{r}\bar{\text{e}}\text{l}}\mathbf{h}_{\tau,\text{flat,r}\bar{\text{e}}\text{l}} + \mathbf{n}_\tau, \tag{3.25}$$

where $\Phi_{\tau,\text{flat,rel}}$ and $\Phi_{\tau,\text{flat,}\bar{\text{rel}}}$ are matrices including the column vectors of $\Phi_{\tau,\text{flat}}$ that are associated with the relevant and non relevant channel components respectively.

The second term of (3.25) can now be viewed as a second noise term. Under the assumption that the non relevant channel components are weak as compared to the strong channel components, a reasonable channel estimate can be constructed using the pseudo inverse (denoted by †) of the pilot submatrix associated with the relevant channel components

$$\hat{\mathbf{h}}_{\tau,\text{rel}} = \Phi^\dagger_{\tau,\text{flat,rel}} \mathbf{y}_\tau. \tag{3.26}$$

The pseudo inverse can be calculated, as the matrix $\Phi_{\tau,\text{flat}}$ in (3.24) has been designed so that the column vector of any submatrix $\Phi_{\tau,\text{flat,rel}}$ with ≤ 6 columns has full rank.

In the example above, no correlations are taken into account. However, the Kalman filter application in Paper IV takes correlation in time, frequency and space into account through the use of the state space model (3.1). When the temporal correlation is considered it can to some extent compensate for the lack of pilots in order to estimate more than K channel components. However, an investigation of the quality of such estimations is outside the scope of this thesis.

If the channel on the different pilot bearing resources are not flat fading, then the channel vector $\mathbf{h}_\tau$ will include one element per pilot bearing resource and antenna port, as in the examples in Section 1.2.1.

The matrix in (3.24) provides an example of how a pilot matrix may be designed using the modulation alphabet $\{-1,0,1\}$ such that any number of up to K different channel components can be selected while ensuring that the pilot matrix $\Phi_{\tau,\text{flat,rel}}$ has full rank. As the number of channel components N increases, the task of finding such a matrix may become increasingly difficult unless the pilots are allowed to take on more values. In Paper IV, the pilots sent over the n'th channel on the k'th resource takes on complex values $e^{j\theta_{k,n}}$ where the phase values $\theta_{n,k}$ are found off-line to ensure that most submatrices $\Phi_{\tau,\text{flat,rel}}$ that consists of a number of $\leq K$ of the column vectors in Φ_τ have a low condition number. That in turn means that regardless of which channel components are the strongest for a user - and by extension regardless of where in the system a user is positioned - it will be able to estimate its own relevant channel components.

To further ensure that the pilots support as many of the potential users as possible, cycling pilots are used. This means that a set of different pilot matrices $\Phi_{\tau,\text{flat}}$ are defined, and applied in a cyclic (repetetive) pattern. It is a concept that was suggested for channel estimation based on uplink pilots in [45] in order to increase orthogonality between user pilots. In Paper IV the concept of cycling pilots is used to make sure that as many potential users as possible will be able to estimate their relevant channel components.

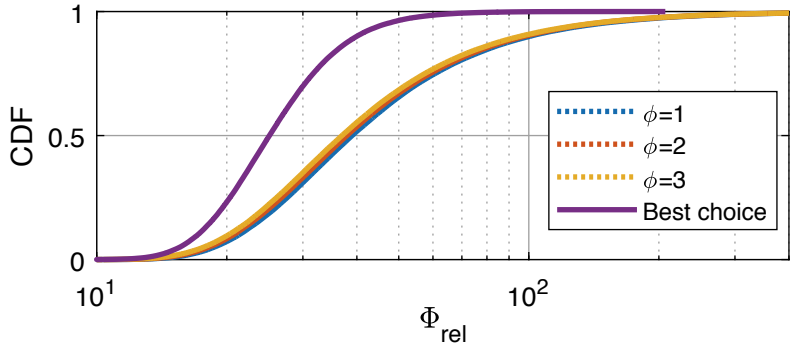

Figure 3.7. CDFs of the condition number of the pilot submatrix $\Phi_{\tau,\text{rel}}$ for different sets of 18 relevant channel components (i.e. different user positions) for three different pilot matrices $\Phi_{\tau,\text{flat}}$ defined by a parameter ϕ. Here, $K = 18$ and $N = 72$. The line "Best choice" shows the CDF of the condition number that will be given if, for any submatrix, all three choices can be considered and the best can be chosen. Further details are given in Paper IV.

Figure 3.7 shows the distribution of condition numbers among the submatrices of three different pilot matrices $\Phi_{\tau,\text{flat}}$ when $K = 18$, $N = 72$ and the phase angles $\theta_{k,n}$ of the pilot signals $e^{j\theta_{k,n}}$ are defined by a parameter ϕ through

$$\theta_{k,n} = (k\phi)^n. \tag{3.27}$$

All these three pilot matrices provide low condition numbers for most of the submatrices $\Phi_{\tau,\text{flat,rel}}$, hence most of the potential users will be able to estimate up until 18 relevant channel components, by using (3.26). However, there are some submatrices that have poorer condition numbers.

If instead of selecting only one of these pilot matrices, all three are used in a cycle, such that $\Phi_{\tau+3,\text{flat}} = \Phi_{\tau,\text{flat}}$ then at least one out of every three times will give a well conditioned submatrix. If temporal correlation is used, e.g. through a Kalman filter, then these well conditioned matrices can be used to improve the estimates also for the other times. If temporal correlation cannot be used, e.g. if there has not been any time to gain knowledge of the channels temporal behaviour, then at least every third set of pilot measurements can be used.

Including cycling pilots, means that the Kalman filter will become cyclo stationary with $\mathbf{P}_{\tau+c} = \mathbf{P}_\tau$ as $\tau \to \infty$ where c is the lenght of one cycle. The error covarianc matrices and Kalman gains of the cyclostationary Kalman filter can be found by solving an extented Riccati equation, see [1] for details. This will increase the off-line complexity of the Kalman filter, but does not affect the on-line complexity.

3.5.2 Joint estimation of channels at adjacent subcarriers

The accuracy of the estimates obtained by a Kalman filer can be improved by including channels for many correlated adjacent subcarriers in the channel vector $\mathbf{h}_\tau$. The more subcarriers that are included, the more accurate the estimate will be, if the noise components in (3.2) on different subcarriers are uncorrelated. The Kalman filter can then reduce the noise influence by averaging over subcarriers. However, estimating several subcarriers jointly comes at the price of added complexity. The benefit of including more subcarriers decreases with the number of subcarriers added and is depending on the finite coherence bandwidth of the channel. If the coherence bandwidth is small, then the added bonus of including many subcarriers will be small.

A way to avoid some of the extra complexity associated with including many subcarriers is to use a post processing step that performs averaging of channel estimates over subcarriers, for example by using a one dimensional Wiener filter that utilize the frequency correlation. The input to this post processing can be estimates that have been produced by a set of low complexity parallel Kalman filters each of which only includes a few subcarriers.

In the case of downlinks of FDD massive MIMO that use the coded pilots introduced in Section 3.5.1, it is important that the Kalman filters span the number of subcarriers that are included within one pilot code block. Then other measures may be needed to keep the complexity down. In Paper IV this is handled by regarding the second term in (3.25) as a noise term. The noise is then no longer white or Gaussian, which does affect the optimality of the Kalman filter.

3.5.3 Where to position the filters and the feedback overhead in FDD systems

Channel estimators for FDD downlinks may be physically located at the terminals (users) or on the network side. In the former case, the estimates or predictions are fed back to the network over uplink control channels. In the latter case, measurements need to be fed back, and the predictors operate based on these measurements. The feedback load per predicted resource block per user depends on the detailed system design used in these two alternatives. There are different advantages to each of these alternatives. These are discussed in Section 4.6 of Paper I.

An investigation of where to place the estimator in a CoMP JT setting was performed in [46]. The authors found that in a centralized setup, with a central unit calculating the joint beamformers, the placement did not affect the gains. However, in a distributed scenario, it was found in [46] that gains were increased if predictions were performed on the network side.

A benefit of locating the Kalman filters at the network side is that this often allows for more sophisticated equipment, so a higher computational complex-

ity might be allowed and so the number of jointly predicted subchannels can be set high. In addition, the prediction quality of the users will vary less as the difference in the different user equipments will affect the predictions less. If would also allow the network side to easily adapt the prediction horizon.

On the other hand, if the channel estimators are located at the user side, adaptive quantization schemes based on filtering or prediction performance may be used. Then, if a user feeds back e.g. two predicted channels where one of them is very accurate and the other is very inaccurate, the more accurate channel may be quantized with higher granularity than the inaccurate channel. Methods for compressing feedback data can be studied in e.g. [47, 48, 49]. If the channel is sparse in the time domain, then the reporting of only the significant time domain taps of the channel, as suggested in [48], will be more advantageous than source coding the frequency selective frequency domain channels. In [49], the explicit feedback is evaluated in a similar manner using compressed sensing, demonstrating its potential for sparse channel impulse responses. In [47] early methods for compressing feedback were compared.

Moreover, scheduling and link adaptation decisions are today based on quite coarse CQI. As the number of users considered for scheduling is often much larger than the number of users that is actually scheduled, it would cause a large feedback overhead to feed back pilot measurements for all of the potential users. If the channel estimators are located at the user side, and if a small extra reporting delay is acceptable between scheduling and transmission, then the users may report complex channel gains only for the users/data streams that have actually been scheduled. A similar concept has been proposed in [50].

To keep the feedback overhead low is especially important in a massive MIMO FDD system, as reporting back a large number of channels will add to the already large overhead requirement. The grid of beams concept is important here as it will ensure that many channel components (the non relevant channels) do not need to be fed back.

4. Applications

MIMO downlink transmission techniques that serve multiple users are a part of LTE 4G [8] and are becoming increasingly important in the study of future systems. As MIMO transmit schemes rely on CSI at the transmitter [13, 51], channel estimates of sufficient accuracy become crucial. This chapter will focus on different MIMO downlink transmission scenarios that are candidate techniques for 5G systems.

MIMO transmission to one user improves the link throughput, with a gain in spectral efficiency that depends on the number of receiver antennas at the user equipment and on the effective rank of the channel matrix. A large additional part of the system gain of MIMO is associated with serving multiple users simultaneously on the same resources. This is because the beamforming gain that can be achieved from MIMO in an ideal case for one receiver with fixed number of receiver antennas, with perfect CSI and perfect hardware, increases the capacity logarithmically with the number of added transmit antennas. With imperfect CSI and/or hardware the gain will saturate. By adding more users when extra transmit antennas are added, each user may not receive increased capacity but the sum-capacity can be increased linearly with the number of antennas added. Multiuser diversity was first introduced in [52]. Since then, the potential for multiuser scheduling gain has been thoroughly investigated for single cell Single-Input Single-Output (SISO) transmission and for single cell MIMO transmission, see e.g. [23, 24].

For the past decade, evolution of multiuser MIMO has moved in two main directions: massive MIMO [23, 53, 54, 55, 56] and CoMP JT (also known as network MIMO) and coordinated beamforming [57, 58, 59]. Each of these, and combinations of them, have been identified as key enablers for the fifth generation mobile system [39, 60, 61, 62]. The channel estimation challenges of downlinks in such scenarios, with a very large number of antennas and radio channels, have motivated the work presented below.

4.1 Channel estimates with non orthogonal pilots for massive MIMO FDD systems

In massive MIMO downlinks the number of transmit antennas at the base station is very large, which allows for very narrow beamforming during transmission. This allows the system to potentially serve a very large number of terminals (user equipments) within the same resources. Serving multiple users

will increase spectral efficiency, as long as the added data rate to a new user is larger than the reduction in rates to other users caused by allowing an additional user to share the transmission resources.

It has been shown that, provided that an adequate CSI quality can be maintained, in an ideal scenario massive MIMO can provide a linear increase in beamforming gain with the number of transmit antennas [54]. That is, the signal power in the Signal to Interference and Noise Ratio (SINR) increases linearly with each added antenna. However, the power of the interference caused by the simultaneous transmission to other users depends on the cross correlation between channels to the users that are scheduled to be jointly served.

If the users are few compared to the number of serving antennas at the base station, then the channel vectors to each user (which consists of the channels between that user and each transmit antenna) will be almost orthogonal with high probability to the channel vectors of the other served users. In such a scenario the user can be served without the need of interference mitigation. Then, MRC can with high probability be used to increase the SINR linearly with the number of added transmit antennas.

Although serving a small number of users compared to the number of transmit antennas at the base station may be beneficial to avoid interference, it is not necessarily the best in terms of maximizing the sum throughput. In fact, due to the logarithmic relationship between SINR and capacity, it may be benificial to allow more interference (and hence lower SINR) in order to be able to serve more users. An investigation of how many users to serve can be found in e.g. [63].

An issue with massive MIMO in FDD systems, which has been pointed out by [54] among others, is that it may come at the cost of a very large pilot overhead. In order to secure accurate CSI, pilots are required to be sent regular intervals, and the most accurate channel estimates are provided by resource orthogonal pilots, as shown in e.g. [64, 44]. However, in an FDD system pilots must be transmitted in the downlink over each channel that is to be used for the data transmission. The large number of antenna elements at the base station would then cause a large overhead. As an example, assume that resource orthogonal pilots should be transmitted every 5 ms and every 180 kHz, corresponding to every 70'th OFDM symbol and every twelfth subcarrier in a 4G system [8], from a massive MIMO base station. Then if the base station had e.g 256 antenna elements, the overhead would be 30%, and it would increase linearly with each added antenna at the base station.

This problem has led to a focus on TDD systems when it comes to massive MIMO. In a TDD system, the scheduled users can transmit pilots in the uplink, and channel reciprocity can be used to gain estimates of the downlink channels based on the uplink pilot measurements. Although TDD systems are also limited, due to e.g. imperfections in channel reciprocity, limited transmit power at the user, hardware impairments and lack of downlink interference estimates [55, 62, 63, 65], they are a sound choice for implementation of mas-

sive MIMO. For example, a recent measurement based investigation in Lund, Sweden provided results in favour of TDD [66].

So the question arises: Why should we investigate massive MIMO in FDD systems? The most important argument is that a large part of the spectrum is presently allocated to FDD and will probably remain so for many years to come. It would be unfortunate not to be able to take advantage of the potential massive MIMO gains in these spectral resources. In [56], the authors identified enabling massive MIMO for FDD systems as the "critical question" for future research on the topic of massive MIMO. Solving the joint problem of pilot design and channel estimation for massive MIMO in FDD systems would also allow backward compatibility, which is a desirable feature for next generation systems [67].

To be able to use massive MIMO in an FDD system, it is important that the channels needed for adjusting the transmission scheme can be estimated without introducing a large overhead, meaning that the use of non orthogonal pilots becomes necessary. The solution should also be such that it supports a large set of users with very different channel vectors, such that any user in the system can be able to use pilot based measurements to estimate its relevant channels.

As an additional more ambitious requirement, the work described in Paper IV has added that the pilot and channel estimation scheme should also enable joint estimation of channels from multiple base stations within a cooperation cluster to support CoMP JT. In such a scenario, the total number of transmit antennas becomes even larger, further complicating the use of orthogonal pilots for each antenna within the cluster. In Paper IV, such a scenario is assumed, with a set of FDD base stations, each with a massive MIMO antenna. Such a problem becomes so challenging that it is hard to find a single design principle, or algorithm, for solving it. Paper IV instead proposes to use a set of tools and designs that work in synergy, and together provide a solution to the problem.

As a first step, in Paper IV, a fixed grid of beams which has been proposed in e.g. [68, 5, 69], is formed at each base station. The main reason for this is that channels to a user from different antennas that are co-located will have similar statistics. This will make it hard to separate them at the user when non orthogonal pilots are used. By introducing a fixed grid of beams, where each beam is controlled by an antenna port, the signal energy from each antenna port is directed in different directions. Therefore they will experience different shadow fading. From the perspective of a single user, some beams will be strong and some will be weak [70]. The user then only has to estimates its strongest channels, which will be referred to as its relevant channel components.

The second step is to introduce coded (and non orthogonal) reference signals. The N beams within the cluster of coordinated base stations then send different sets of pilots symbols in a set (pilot word) of K resources elements.

The sets of N K-symbol words are the coded reference signals, which are known to all users. In order to ensure that all users will be able to estimate their relevant channel components, a vector containing the pilots transmitted over K pilot bearing resources from one antenna port should be linearly independent to any set of vectors that contains the pilots transmitted by up to $K-1$ of the other antenna ports. Then, any user will be able to estimate up to K relevant channel components by simple matrix inversion. An example of such a pilot code was given in Section 3.5.1.

Although the above ensures that a matrix consisting of the pilot vectors of up to K relevant channel components, denoted relevant pilot matrix, has full rank, and is hence invertible, it does not ensure that all potential relevant pilot matrices are well conditioned. Inversion of a poorly conditioned matrix may cause small errors to have large significance. Pilot codes should therefore be chosen off-line to ensure that as many potential sets of relevant channel components as possible provide well-conditioned matrices.

For this reason, as a third step, the scheme investigated here utilizes cyclic pilots, ensuring that the users that have relevant channel matrices with high condition number will change over time and therefore all users should be able to estimate their relevant channel components, by using some of the cyclically changing pilot patterns.

As the fourth step, correlation in space, time and frequency among the channel components is used to improve channel estimates, compared to a simple matrix inverse solution. The most advanced of the channel estimation schemes investigated in Paper IV is the Kalman filter which utilizes all of these correlations. However, such a Kalman filter for multiple massive MIMO downlinks will have high complexity if all channel components are to be estimated, and therefore a reduced Kalman filter, which estimates only the relevant channel components for each user, is investigated as an alternative.

4.1.1 Relations to previous results

Other works that have looked into channel estimation for massive MIMO in FDD include [71, 72, 73, 74, 75, 76, 77, 78]. Similar to the design of Paper IV, these works assume non orthogonal pilots and utilize some type of correlation to improve the estimates.

In the earliest of the works above, namely [71, 72], user specific pilot design was suggested. Based on downlink transmission of these pilots, the terminal generated an uplink feedback to the base station which then utilized Kalman filters to estimate CSI. Another single user scheme, partially based on the use of compressed sensing, is proposed in [73]. These concepts would demand user specific pilot resources, so the overhead increases with the number of users and the benefit achieved by using non orthogonal pilots decreases rapidly as the number of users increase.

Authors of [74, 75, 76, 77] instead optimize the pilots off-line to improve the average channel estimation for the scheduled users. The required pilot overhead would in [74, 75], increase with the number of active users. The work of [76] assumes sparsity in the Channel Impulse Response (CIR) and correlation between the channels from different antennas. This estimation scheme does not provide gains when CIR are not sparse, which often occurs in real channels [14].

These multiuser methods, though they are an improvement compared to the single user case, still require the pilot pattern to be optimized each time new users are scheduled. The solution presented here is based on the fixed grid of beams and cycling sets of pilots, which do not have to be optimized for the set of scheduled users and their channels. The combination of a fixed grid of beams and cycling between fixed sets of coded pilot vectors will ensure that most *potential* users can estimate their channels. Then, optimized pilots need not be fed back and even users not yet scheduled for service can prepare for transmission by estimating their channels based on the available downlink pilots.

A somewhat related idea is proposed in [78] where pilots are transmitted over a number of beams, lower than the number of transmit antennas at the base station. This work focuses on estimating the strongest one or two beams, claiming that this is sufficient to get close to full sumrate capacity gain. While this may be reasonable for MRC transmission with high SNR and few users (so that inter user interference can be ignored) it will not be adequate when interference mitigation is necessary.

A fixed grid of beams, similar to that in [68, 5, 69] was investigated in [66]. The authors found that for clusters of closely spaced users, or a hot spot scenario, the performance with the grid of beam concept is reduced. This problem is caused by using a fixed allocation of a low to moderate number of beams that are designed to span the full cell area. The number of beams serving a hot spot area might then become small and inadequate. A potential remedy for that is to rearrange the fixed grid of beam on a slower time scale as suggested in [69].

4.1.2 System design

The general proposed design is illustrated in Figure 4.2. Here, an OFDM FDD system has a total of N_{PRC} physical downlink radio channels between the serving antennas and each single antenna user. The serving antennas may all be set at the same location or, as in the simulations in Paper IV, distributed among the base stations included within a CoMP cluster.

Many pilot symbol sets are possible. In the design investigated in Paper IV the antennas are controlled by N_{CC} antenna ports, each controlling a separate beam in the fixed grid of beams.

Here, the word fixed referers to fixed over a long time period, e.g. several seconds. However, the fixed grid of beams can change whenever the distribution of users changes significantly, e.g. if an office building is empty during night then the grid of beams can be adjusted such that it transmits little or no energy into that building.

During the pilot transmission phase, each beam transmits a set of pilots over K available downlink pilot bearing resources. This transmission is repeated in time, with a frequency that in appropriate with respect to the mobilty of the intended users. The pilot symbol that is transmitted by the n'th antenna port on the k'th resource a time τ is given by

$$\varphi_\tau(k,n) = \exp(j\theta_\tau(k,n)), \tag{4.1}$$

where $(\theta_\tau(k,n)$ is a phase angle. Thus, all complex valued pilots are designed to have the same power.

Based on pilot measurements (3.25) the relevant channels are estimated, either directly at each of the M users, or at the base station based on fed back measurements. If the users estimate the channel, then the CSI of the relevant channel components is fed back, as depicted in Figure 4.1. In a second step, the data transmission phase, the CSI is used to design a precoding matrix for the antenna ports, e.g. through MRC.

The phase angles θ_τ of (4.1) need to be adjusted such that the pilots transmitted over any set of up to K relevant channel components will form a relevant pilot matrix $\Phi_{\tau,\text{rel}}$ with full rank, as explained in Section 3.5.1. For some parameters ϕ_τ this can be achieved by setting the phase angles according to (3.27). The real-valued scalar ϕ_τ is a design parameter that should be chosen off-line to ensure that the requirement of full rank is obtained.

For the coding by (4.1) and (3.27), cyclic pilots, as described in Section 3.5.1 are implemented. These are then cycled with a period μ over time such that $\phi_\tau = \phi_{\tau+\mu}$. It is then likely that any subset of relevant channels will receive a well conditioned relevant pilot matrix for at least one of the cycling pilot codes. Over a time period of μ, the user will have at least one good estimate, and a number of reasonably good estimates. If temporal correlation is accounted for, then this good estimate can also be used to improve all estimates.

Note that introducing cycling pilots does not introduce any additional overhead, nor does it require that all users are equipped with estimators that can utilize temporal correlation. Channel estimators that utilize the temporal channel correlation will be able to improve their estimate by combining estimates obtained with different subsequent pilots code vectors, thereby reducing the influence of badly conditioned cases. In particular, the Kalman filter (3.4)will downweight the influence of measurements $\mathbf{y}_\tau$ for which the pilot matrix Φ_τ is badly conditioned. Other estimators can still use the pilots at each time.

In addition to the non orthogonal superposed pilots that are transmitted frequently, the work in Paper IV also suggests that resource orthogonal pilots are transmitted over all beams sparsely, e.g. every 0.5 s, see Figure 4.2. As

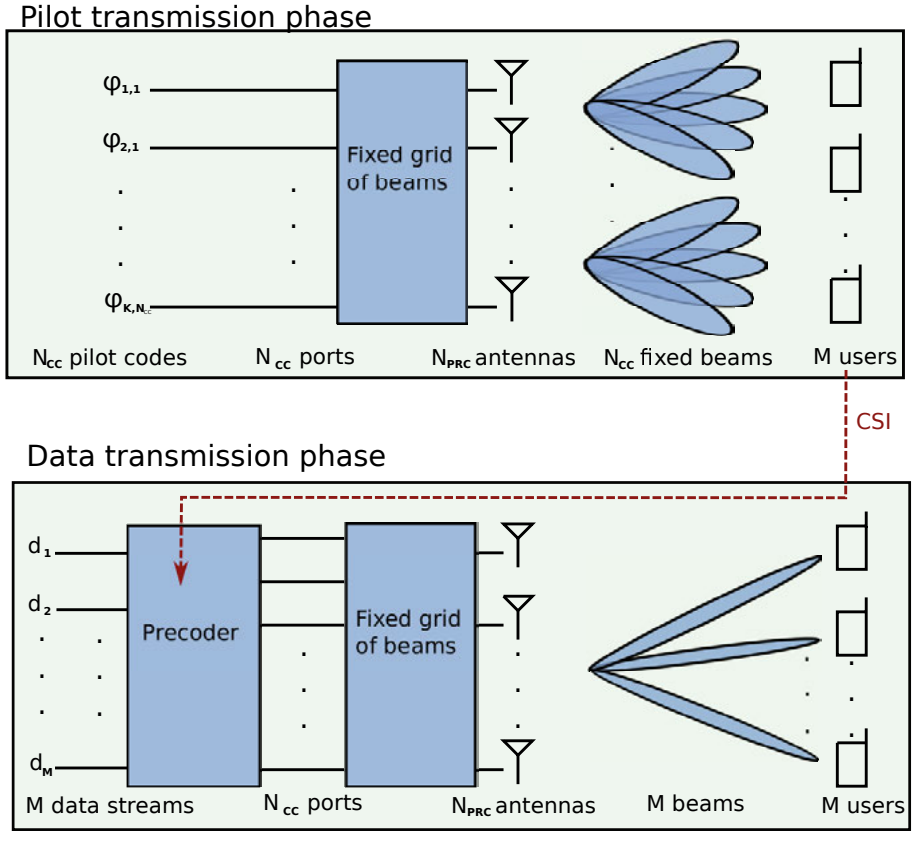

Figure 4.1. Pilots are transmitted in the downlink over a fixed grid of N_{CC} beams controlled by N_{CC} antenna ports. Users estimate subsets of the channel components and feed back these estimates over the uplink. During the data transmission phase the N_{CC} beams are precoded, e.g. by MRC or interference mitigation precoding, to direct the signal energy to each user, and to potentially remove interference from the other users.

shadow fading only changes on a long time scale, of at least several of hundreds of ms for pedestrian users, these sets of resource orthogonal pilots can be used to estimate the channel correlation over space and frequency. It is also used to find the set of relevant channels for new users that enter the system and to detect changes in the sets of relevant channel components for an existing user due to shadow fading.

As these orthogonal pilots are repeated infrequently, they do not introduce a large extra overhead cost.

Based on the measurements (3.25), the relevant channels are estimated either through LLMSE estimation or by a reduced Kalman filter. The LLMSE solution is given by

$$\hat{\mathbf{h}}_{\mathrm{rel},\tau} = \mathbf{R}_{h,\mathrm{rel},y}\mathbf{R}_{y}^{-1}\mathbf{y}_{\tau}, \tag{4.2}$$

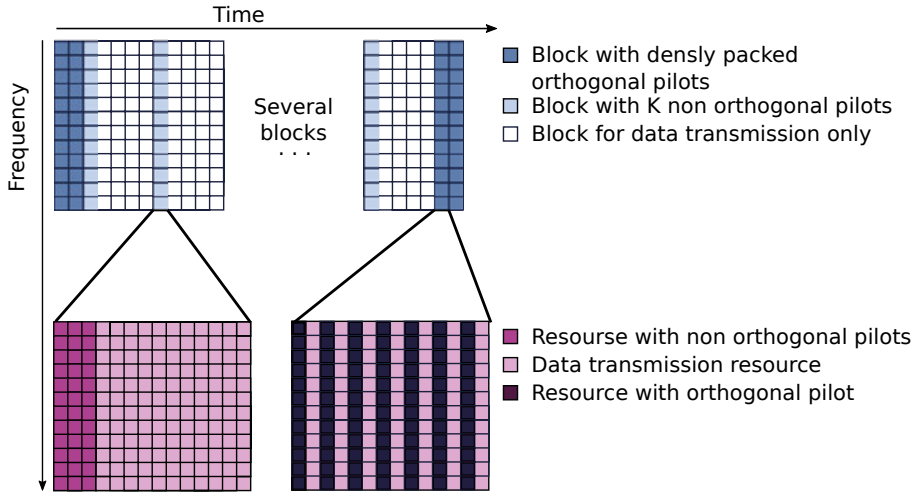

Figure 4.2. Exempel of pilot structure in the downlink. Resources are divided into blocks of 12 subcarriers and 14 OFDM-symbols (as in LTE). Every fifth block contains K pilot bearing resources that can be used for channel estimation. On a slow time scale (every 500:th block), two subsequent blocks include 50% of the time-frequency resource slots for fully orthogonal pilots. The total pilot overhead in this example is approximately 4.5%.

where $\mathbf{R}_{h,\mathrm{rel},y} = E[\mathbf{h}_{\mathrm{rel}}\mathbf{y}_\tau^*]$ is the cross covariance matrix between the vector of relevant channel components and the measurement signal and $\mathbf{R}_y = E[\mathbf{y}_\tau\mathbf{y}_\tau^*]$ is the covariance matrix of the measurement signal. The details on how to estimate these are given in Paper IV.

The LLMSE solution is attractive for user equipment that require bursty data. The user would then be able to listen to the orthogonal, sparse pilots and then, whenever it needs data, be able to quickly estimate its strongest channel components, without needing to wait to gather information about the channel's temporal correlation.

For the reduced Kalman filter estimate, the covariance matrix of the measurement noise is adjusted to model both the interference term $\Phi_{\tau,\bar{\mathrm{rel}}}\mathbf{h}_{\tau,\bar{\mathrm{rel}}}$ and the noise term $\mathbf{n}_\tau$ of (3.25) as measurement noise. Then, the filter equations are solved as described in Chapter 3. For details, please see Paper IV.

4.1.3 Results and conclusions

To validate our concept we set up a system level simulation using the Matlab based, open source, Quadriga channel simulator, developed by the Fraunhofer Heinrich Hertz institute [20, 6]. A NLOS scenario was considered with nine base stations, each equipped with 32 transmit antennas. In total 288 antennas transmit non orthogonal pilots over $K = 18$ pilot bearing resource by $N_{CC} = 72$ beams over fading channels to 100 pedestrian users. The performances evalu-

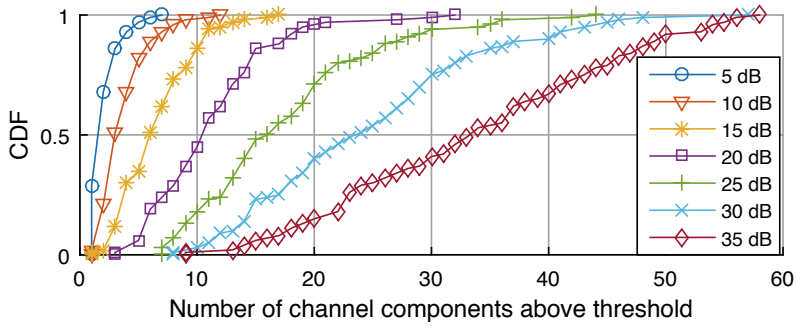

Figure 4.3. CDF of the number of relevant channel components at different user positions when using different thresholds in dB, relative to the power of the strongest channel.

ation was thus performed for a case where the pilot overhead was reduced by a fraction of $18/72 = 0.28$ as compared to a case with orthogonal pilots on all 72 beams. Compared to a case with individual orthogonal pilots per antenna, the overhead reduction is $18/288 = 0.063$. For more details on simulation parameters, please see Paper IV.

Relevant channel components

How many relevant channel components that are necessary to estimate depends on how the data is to be transmitted. For example, in [2], it was shown that to provide good CoMP performance through joint coherent interference mitigation, channel components that were in the range of 20 to 25 dB below the strongest channel components should be included and be used by the linear precoding algorithm.

It is therefore important to know how many beams will be in this range. Figure 4.3 shows the CDF of the number of channel components that would be relevant if a transmit scheme utilizes only channel components with power above a threshold relative to the the strongest channel component in the investigated simulation environment.

Assuming a threshold of 20 dB, results in Figure 4.3 indicate that no more than 15-20 channel components (out of a total of 72) would then typically need to be estimated by a user.

Estimations Performance

Figure 4.4 shows the filter estimate NMSE as a function of the channel component number in terms of power, averaged over subcarriers and users when sixteen relevant channel components are estimated through an LLMSE filter and a reduced Kalman filter. As a comparison, it is also investigated what happens when no correlations are accounted for, i.e. when the channels are estimated by simple inversion as in (3.26).

In addition, results are shown where Kalman filters have been used assuming resources orthogonal pilots. In this set-up each channel component is then assigned resource orthogonal pilots for all available subcarriers and for one out of 72 subsequent OFDM symbols. The total pilot power budget is the same for the cases with resource orthogonal pilots as for the non-orthogonal pilots. Channel estimation by inversion by (3.26) is also performed for the orthogonal pilot case. In all case, the SNR per channel component is the inverse of the NMSE achieved by inversion with orthogonal pilots (the thin solid line in Figure 4.4).

While channel estimation by (3.26) may be sufficient for the strongest channel components, it quickly degrades for weaker ones. Comparing channel estimation by (3.26) to the LLMSE filter, we see that by utilizing the space and frequency correlations, estimates are greatly improved, especially for the weaker channel components.

A further improvement of approximately 5 dB can be achieved by utilizing temporal correlation by introducing the reduced Kalman filter.

The most effective approaches previously suggested for channel estimation for FDD massive MIMO is to optimize pilots for a specific set of users [74, 75, 76, 77]. To relate the presents results to those obtainable with pilot optimization, we can compare the channel estimates based on non orthogonal pilots to those based on the orthogonal pilots when the reduced Kalman filter is used. In an extreme situation where the union of the sets of relevant channel components of all the scheduled users includes no more than K (here 18) channel components, an optimization of pilots would result in the estimation performance close to the one here obtained with orthogonal pilots [1]. In a more realistic situation, where the union of the sets of relevant channel components increases with an increasing number of scheduled users, the estimation performance would move towards that of reduced Kalman estimation with non orthogonal superposed pilots.

This reasoning indicates that the here proposed flexible solution, which does not require pilots to be re-optimized every time a new user is scheduled, should in the worst case scenario result in a 5 dB reduction of the estimation NMSE performance, as compared to a case with optimized pilots for multiple users. However, it should be stressed that a direct comparison with the pilot optimization and channel estimation schemes of [74, 75, 76, 77] has not yet been performed.

There is potential to reduce the estimation error further, by estimating also additional channel components that are below the threshold of relevance. This is illustrated by Figure 6 of Paper IV. It compares the estimation performance

[1]For a scenario of one of very few users, then it is likely that performance can be made much better with optimized pilots as a result of optimizied power reallocation. With multiple users and multiple base stations, it is unlikely that this will provide large gain. Reallocation can then mainly remove power from beams that are of little value to any of the users. However, a similar effect can be achieved by using the beam deactivation scheme suggested in [69].

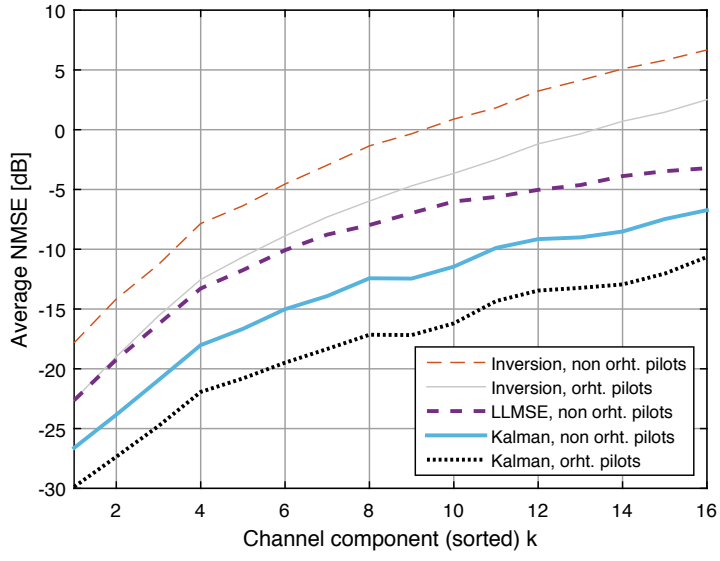

Figure 4.4. Average NMSE sorted after the strength of the channel component.

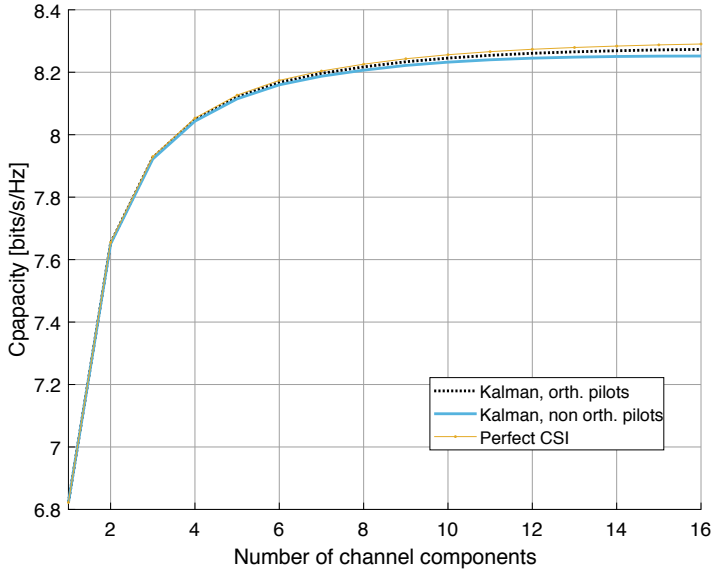

Figure 4.5. An illustration of the single user maximum ratio beamforming gain, based on Shannon capacity, as a function of the included number of channel components. CSI is based on the two Kalman estimates in Figure 4.4. Note that the curves overlap.

of the reduced Kalman filter to that of a Kalman filter which estimates all 72 channel components. These results indicate that a full order Kalman estimation performance with non orthogonal pilots becomes almost as good as that gained by reduced-order Kalman filters with orthogonal pilots. However, as the Kalman filter complexity grows with the square of the number of channel components, the estimation of all channel components becomes highly computational complex. Therefore these results have only been evaluated for one of the potential user positions, i.e. for only one potential set of relevant channel components, and should hence be interpreted with care.

Paper IV also discusses potential methods of improving the reduced Kalman filter to attain a performance closer to that of the full-order filter without increasing complexity to unrealistic levels, e.g. by tracking the correlation properties over time of the interference term $\Phi_{\tau,\bar{\text{rel}}}\mathbf{h}_{\tau,\bar{\text{rel}}}$ in (3.25) through first or second order AR models. Such investigations are left for future work.

Capacity of MRC beamforming using estimated channels

The impact of the estimation errors in terms of MRC beamforming gain is here illustrated in Figure 4.5. Maximum ratio beamforming to one user is performed jointly by the nine base stations, by using a joint precoder, see Figure 4.1. For each investigated user position, the precoder utilizes the estimated channels of the K strongest fixed beams for a particular user. Figure 4.5 shows how the resulting spectral efficiency increases with the number of K beams that are allowed to be utilized, on average over 100 user positions. The results in this figure are assuming that the estimation performance is given by the Kalman estimates presented in Figure 4.4. The Shannon capacity shown in the figure is for the single user case.

From Figure 4.4 we see that there is next to no beamforming gain from having orthogonal pilots in this particular scenario. The reasons for this are as follows: First, the estimation quality is already very good for the strongest channel components and an extra gain in accuracy will only translate into a very small capacity gain. Second, MRC beamforming gains are robust to estimation errors. Third, as capacity grows logarithmically with SNR, adding extra channel component, that have low SNR as compared to the strongest channel component, to the beam provides very little extra gain; note the saturation of the curves in Figure 4.5.

A separate study of interference supression by regularized zero forcing precoding in a CoMP cluster with massive MIMO antennas under a similar simulation set-up as here, but with 288 beams, is presented in [69]. This study also includes a beam deactivation algorithm to ensure that no power would be wasted on beams that are weak to all users within the system. The results presented there, with 10 % pilot overhead, show that performance loss, in terms of payload spectral efficiency, was approximately 12% compared to when perfect CSI was used.

4.2 Channel prediction for coordinated multipoint joint transmission

In a traditional cellular network, each base station serves the users that are located within its own cell, see Figure 4.6. When a user moves from one cell to another, the base stations cooperate in the *hand-over* procedure, but otherwise they serve their users independently of each other. The base stations can use different resources and/or different beams to avoid *intracell interference*, i.e. that the energy leaks between the different users' messages. However, as base stations generally do not cooperate, except for in the hand-over procedure, energy might leak between cells, causing *intercell interference*. This interference decreases the data throughput for the users and is especially severe for users close to the cell edges.

One option to decrease the intercell interference from nearby base stations is to partition the available frequency spectra such that each base station only uses a fraction of the available spectrum. This method is called frequency reuse. However, frequency reuse lowers the capacity of the system, as the base stations will only be able to serve it's users with a fraction of the spectrum.

To increase spectral efficiency, present 4G systems allow all base stations to utilize the whole available spectrum (this is called frequency reuse 1). The disadvantage of doing so is that the signals from several base stations will in general be quite strong at the cell edges, thus causing intercell interference to be the main limitation for users located at the edges.

A further way of increasing user throughput is to increase the spatial density of base station and thus decrease the size of cells. If this is combined with frequency reuse 1, then intercell interference can become an even bigger problem.

The intercell intereference can be limited by allowing a cluster of base stations to cooperate to serve some or all of the users within their cluster of cells, using so called CoMP techniques. These were introduced in the beginning of the millennium as a way to increase the spectral efficiency of downlink transmission [57, 58, 79].

In a CoMP system, base stations share information over links that interconnect the base stations. These might e.g. be radio channels or fiberoptic cables. The base stations may be interconnected in different ways. The direct links between two base stations are referred to as cross links. Links may connect base stations to a CU which may be responsible for the joint transmission decisions. For future wireless networks, it is likely that many more base station nodes will be deployed. With more nodes, operators may want these to be less advanced (and thereby less expensive), perhaps simply consisting of remote radioheads. In such scenarios, a large part of the data and signal processing may be moved from the base stations into Radio Access Network (RAN) clouds that may include centralized baseband units. The links between radio access point and a centralized processing unit is in the literature often called

fronthaul link, if some or all of the RAN functionality is centralized [80]. In this section and in Paper II and Paper III, the name backhaul links will be used for all the connections in the fixed network side that are used to enable the CoMP processing.

Downlink CoMP is often divided into two categories, 1) JT and 2) JS and/or JB [81, 82, 83, 84]. In the first category, multiple base stations attempt to transmit to one, or more, users simultaneously, see Figure 4.7, while in the second category, base stations coordinate their transmission, e.g. such that they avoid serving closely located users within the same resources and thereby lower the intercell interference. In JT we try to utilize the potentially strongly interfering links from neighbouring base stations, turning a potantial problem into an advantage. With JS and JB, we mainly strive to avoid interference. For this reason, JS and JB provide lower gains in spectral efficiency than JT, but are more robust to errors in CSI [3, 85].

In CoMP JT, base stations share both payload data and CSI over the backhaul links. Based on the CSI, some or all of the base stations may then be selected to transmit payload to one user, without any pre-compensation of the message symbol intended for that user. As the message is transmitted with radio waves these may sometimes add up constructively at the users, providing a stronger receive signal than if only one of the base stations transmitted. However, when no pre-compensation of the message is used, then the signals may also add up destructively, lowering the total received power[2]. This is called non coherent CoMP JT.

An alternative is that the base stations precode the message symbol, based on the CSI, before transmission, to ensure that the signals from the different base stations add up constructively. This scheme, which is called coherent CoMP JT or network MIMO, allows base stations to serve multiple users on the same time-and-frequency slot, or resource. The messages are then precoded such that, at each user, only the message intended for that user is constructively added, while the messages intended for the other users are added destructively.

The results in this section are extracted from Papers II and III. They aim to investigate

1. How well channels can be predicted. A challenge with downlink CoMP are the long system delays. These include the processing time for channel estimation, the feedback of measurement or CSI feedback from the users to the base stations in FDD systems, the CSI sharing between the base stations over backhaul, the process time for the precoder design and the sharing of the precoder weights between base stations or between a centralized unit and radio access points over fronthaul links in a cloud RAN architecture. The sum of these system delays can be in the order of of tens of milliseconds, which for mobile users will cause the

[2]On average the received power will be stronger.

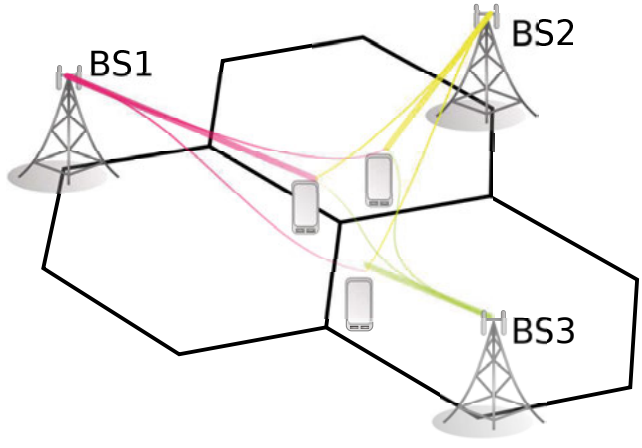

Figure 4.6. A set-up of a cellular network where each base station (BS) serves the users in its own cell only. Different colors indicate messages intended for different users. In the absence of base station cooperation, some of the energy intended for a user in one cell might leak to a user in another cell, causing interference.

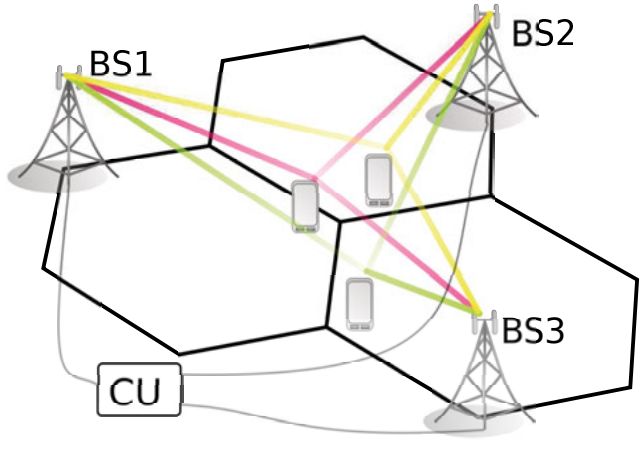

Figure 4.7. A set-up of a CoMP JT scenario where three base stations (BS) transmit jointly to three users. The base stations share information over backhaul links via a control unit (CU).

CSI to be outdated at the time of transmission [86, 87, 88]. The outdating of CSI can to some extent be counteracted by channel predictions [33, 1, 89, 38]. Here, Kalman predictions are investigated for different prediction horizons based on channel measurements.

2. How the performance of coherent CoMP JT is affected if knowledge of the prediction performance (in terms of second order statistics of the prediction errors) is accounted for in the precoding step as compared to when this is not the case. The investigated precoder, which aims to maximize the sumrate, uses not only the Kalman predictions for pedestrian users, but also the knowledge (through error covariance matrices) of their accuracy. It is compared to a precoder that only utilizes the channel predictions.
3. How to deal with different backhaul constraints. Such limitations must be handled by the precoder.
4. How to improve performance by a simple user grouping and resource allocation scheme. Some of the large gains from MIMO techniques are related to serving multiple users simultaneously. However not all users are spatially compatible and a technique for grouping users is required. The proposed scheme assumes that each base station schedules the users within its own cell independently using channel aware scheduling. The users scheduled on a specific resource are then served by CoMP JT.

4.2.1 Background and related work

Coherent CoMP JT has the potential to provide very large gains in spectral efficiency (see e.g. [59, 82, 90, 91, 92, 87]), gains that are especially important for cell edge users [93, 81, 2, 94].

CoMP can also be applied for uplink transmission, either through JS or through joint detection, see e.g. [94, 95]. Joint detection in the uplink is similar to JT in the downlink. However, it is an easier problem in the sense that the processing can be based on fully updated CSI.

The cooperation cluster

There is a trade-off in the number of base stations included in the cooperation cluster. In a cluster with a low number of base stations, the remaining uncompensated intercluster interference will be large, lowering the potential CoMP gains. In a cluster with a large number of base stations, there might be large system delays, causing severe outdating of the CSI. This, in turn lowers potential CoMP gains. Moreover, in very large clusters, base stations will be so far away from each other that this may cause severe synchronization problems.

The problem of clustering has been studied for intracluster interference limited scenarios in e.g. [92, 96, 97, 98]. According to these works, for single antenna base stations, clusters of about three base stations are sufficient in order to achieve most of the CoMP gains. After this point, CoMP gains grow

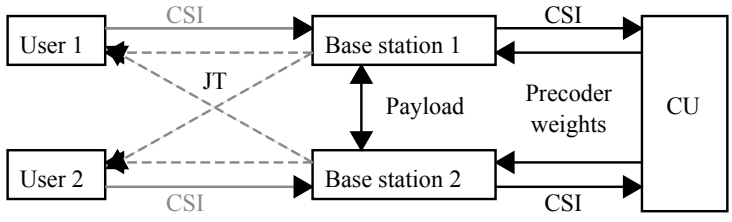

Figure 4.8. Centralized coordination of downlinks in an FDD system. A schematic figure of data transmission of feedback over wireless uplinks (gray arrows) and backhaul links (black arrows). Dashed arrows indicates coherent joint downlink transmission of the payload data.

slowly. For MIMO systems with 2-4 transmit antennas per base station, clusters of 7-9 base stations are required.

Such cluster sizes are fairly small, so intercluster interference levels will still be significant. Therefore external schemes to manage intercluster interference can improve the CoMP gains further. An interesting method to limit the intercluster interference is proposed in [96] and further evaluated in [99]. It uses cluster-specific antenna tilting and power control for this purpose.

Network architecture

A CoMP network may be centralized, distributed or semi-distributed. In a distributed network, each base station forwards CSI to other base stations and then designs a precoder separately. Then, base stations will in general have access to different information, and hence the set of precoders differs from that of a centralized design[3]. In centralized CoMP, all CSI is transmitted to a CU, where the precoder is designed, and all base stations are provided with their relevant scheduling decisions, beamforming weights and payload data.

The CoMP JT system considered in this thesis assumes a centralized architecture as illustrated in Figure 4.7, unless otherwise specified.

For downlink coherent JT in a centralized CoMP architecture, CSI needs to be obtained for all users at their master base stations, i.e. the base station that is the strongest, seen from the user's perspective. These then need to share both the CSI, and the payload data amongst each other over a backhaul network. A CU, which is a logical entity that may be located at one or more of the base stations within the cooperation cluster, then calculates the precoder. Finally, the elements of the precoder matrix are, if necessary, transmitted to the base stations, see Figure 4.8.

A disadvantage of a centralized scheme is that the system delays will be longer as information needs to be transmitted over backhaul/fronthaul links twice. Delays in the fixed network causes outdated CSI, which can severely

[3]For example, if the users feed back CSI to their main base station at frequent intervals, then each base station may have access to less outdated CSI from the users within its own cell than it has to the users within its neighbouring cells.

reduce gains [86, 100, 21]. In order to compensate for the system delays, channel prediction is necessary.

For more information on the different architectures, see [21, 22].

Precoding

The highest gains for coherent CoMP JT are achieved with Dirty Paper Coding (DPC), see e.g. [101]. However, as this non-linear precoding scheme is high in complexity, linear precoding is an important topic of investigations. The primary objective of the precoder is often to limit the intracluster interference. In a system with perfect CSI this can be achieved by channel matrix inversion, which is provided by the zero forcing precoder developed for MIMO, see e.g. [25].

However, there is a risk in using zero forcing for CoMP JT , even in the presence of perfect CSI; Channel gains may from different base stations be very different in amplitude, which is generally not the case for MIMO downlink where all transmit antennas are co-located. The balancing of transmission signal gains that is necessary for channel inversion might then cause the strongest base station's transmit power to become very low as compared with the weakest base station. Although the solution is then still optimal with respect to limiting intracluster interference, the noise and intercluster interference might become large compared to the received signal power, resulting in a poor data rate. An option is then to use an MSE criterion which takes these effects into account as well as the intracluster interference [26].

MSE criteria are attractive as they generally have analytical solutions. However, in practice it is often more useful to optimize over a weighted sumrate criterion, as this is closer to the desired end performance. Such an optimization poses a multidimensional nonconvex optimisation problem.

In the precoder design proposed in Paper II, which is an extension of the desing introduced in [4], a heuristic solution is used, which strives to maximize an approximation of the sumrate by iteratively adjusting a set of parameters, that are criterion weights in a (robust) MSE criterion. The MSE problem stated is derived from a robust linear-quadratic optimal feedforward control framework developed by Öhrn, Ahlen and Sternad in [102]. This general solution is also used for e.g. audio processing, see e.g. [103].

A problem with CoMP JT in a centralized network is, as previously mentioned, that the long delays cause inaccuracies in the CSI. If these inaccuracies are not addressed in the precoder design, then the potential CoMP gains may be lost. Robust precoder design techniques have been suggested for Multiple-Input Single-Output (MISO) systems by [104, 105], and for multi-user MIMO downlinks in [106, 107]. A robust linear precoder for CoMP JT, which is based on an MSE criterion, is suggested in [108].

In comparative studies in [2] and [109], the design discussed here was compared to designs suggested in [108] and to an optimization approach proposed in [110]. It was shown that the here suggested design performed better than

the robust precoder suggested by [108] and close to that of [110], at a much lower computational complexity than the latter.

How to handle limited backhaul

The requirements on backhaul might not always be manageable by the system. Structural constraints, delay constraints, capacity constraints or a combination of these might limit the information that can be transmitted over backhaul. These limitations must be handled by the precoder. In the formulation suggested here, this is achieved by setting some elements in the precoder matrix to zero. These elements represent the input-output connections that are unavailable. Such zeros can be forced into the precoder matrix in different ways.

One option to force zeros into the precoder when backhaul constraints are present is to simply calculate the precoder, based on some criterion, and then set the required elements to zero, see e.g. [111]. However, the resulting precoder is then no longer optimal with respect to the criterion it was optimized for. Another option is to only feed back the channels that are going to be utilized in the transmission and then group users such that the channel matrix becoms block diagonal. Then, through channel inversion, the resulting precoder will automatically have zeros in the required elements [112]. However, such a solution requires JS and may therefore potentially add large demands on backhaul capacity in the user grouping step. Furthermore, the suggested methods in [111, 112] do not consider the inaccuracy of the CSI in the precoding design.

A strength of the here proposed precoder is that the MSE criterion (with tuning of criterion weights to approximate a weighted sumrate solution) can be used also under backhaul constraints. This can be done by introducing large penalty terms in the robust MSE criterion. This modification only requires a low number of extra calculations and is therefore of much lower complexity than methods that use multidimensional searches to optimize the nonzero elements of the precoder, see e.g. [110, 113, 114].

In [115] a different method for forming sparse precoders matrices to fulfil backhaul constraints is suggested. This precoder, which is based on an MSE criterion that includes intracluster interference only, is also low in complexity. However, as the matrix dimensions do not add up in the published design equations of this paper and its simulations are based on very few user positions, it is hard to draw any conclusions from the paper. Moreover, this method does not have the flexibility to optimize over arbitrary criteria, nor does it consider CSI errors in the design.

User grouping and scheduling

A way of determining which users should be jointly served by a CoMP cluster is to form groups based on spatial compatibility, without considering the independent small scale fading of the channels to the users, as in e.g. [111, 116]. In a second step the user groups are then allocated resources. There is a large risk

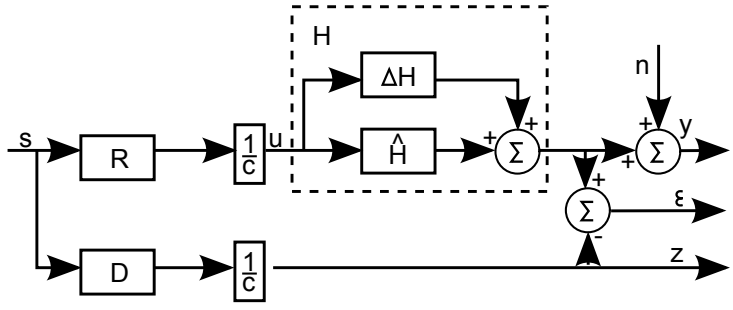

Figure 4.9. System model used for precoder design.

with this type of scheme; the users generally have uncorrelated fading, so one or more of the users in the group may have a poor channel for the particular resource allocated to the group.

The problem with the small scale fading of different users can be solved by jointly performing user grouping and scheduling. However, the combinatorial growth of the complexity associated with finding the best set of users to be grouped into each time-frequency transmission resource makes it infeasible, if the number of transmission resources and users is large. In order to decrease the number of combinations, a greedy user grouping and scheduling algorithm can be used. This has been suggested for downlink single cell MIMO transmission in e.g. [117, 118] and for uplink CoMP in [119].

A user grouping and scheduling algorithm similar to that which will be presented and investigated here can be found in [120], where instead of focusing on increasing rate it is focused on increasing rate dependent utility functions, for the different types of packet streams.

4.2.2 Robust linear precoding design and user grouping

The robust precoder that was introduced in [4] and further developed in Paper II is designed based on the system model shown in Figure 4.9. Here, a vector of complex valued symbols $\mathbf{s} \in \mathbb{C}^{M\times 1}$ to M single antenna users is projected, through a precoding matrix $\mathbf{R} \in \mathbb{C}^{N\times M}$, onto N signals, collected in the vector $\mathbf{u} \in \mathbb{C}^{N\times 1}$, that are to be transmitted from N different antennas. The transmit signals are scaled to ensure that power constraints are met through a real valued scalar $1/c$. This describes the downlink transmission over one narrowband subcarrier in an OFDM system.

The transmit signals are thus assumed to propagate over linear narrowband channels. Furthermore, Gaussian noise is assumed to be added at the receiver. This can be mathematically modelled by

$$\mathbf{y} = \mathbf{H}\mathbf{u} + \mathbf{n}, \tag{4.3}$$

where $\mathbf{y} \in \mathbb{C}^{M\times 1}$ is the vector of received message symbols at the users and $\mathbf{n} \in \mathbb{C}^{M\times 1}$ is the noise and intercluster interference added at each users' receive antenna. The matrix $\mathbf{H} \in \mathbb{C}^{M\times N}$ is the channel matrix containing the complex valued channel elements between each user and the transmit antennas at the base stations. It consists of a sum of a known prediction $\hat{\mathbf{H}}$ and prediction errors $\Delta\mathbf{H}$.

A second system is considered. In this a target vector $\mathbf{z} \in \mathbb{C}^{M\times 1}$ of the desired interference free scaled received symbols is defined by

$$\mathbf{z} = \frac{1}{c}\mathbf{D}\mathbf{s}, \tag{4.4}$$

where $\mathbf{D} \in \mathbb{R}^{M\times M}$ is a diagonal weighting matrix, that ensures that we do not ask more of the system than it is able to carry out. For examples, if all base stations have weak signals to a user, then the diagonal element of that user should be set small, e.g. scaled to the channel gain from nearest base station.

The precoding matrix $\mathbf{R}$ is designed to minimize a scalar weighted robust MSE criterion

$$J = \bar{E}\left[E\left\|\mathbf{V}\boldsymbol{\varepsilon}\right\|^2 + E\left\|\mathbf{S}\mathbf{u}\right\|^2\right]. \tag{4.5}$$

Here $\bar{E}$ denotes the expected value with respect to the prediction errors (the elements of $\Delta\mathbf{H}$) while E denotes the expected value over the involved signals. The weighting matrices $\mathbf{V}$ and $\mathbf{S}$ are design parameters that can be adjusted so that the solution $\mathbf{R}$ that minimizes (4.5) will approximate the solution that optimizes a different objective function, see Paper II for details. In the evaluations presented below, $\mathbf{V} = \mathbf{I}$ and $\mathbf{S}$ is chosen diagonal, with diagonal elements adjusted to maximize an approximation of the sumrate.

Assuming that all message symbols are mutually uncorrelated and also uncorrelated over time, with unit power, the linear precoder that minimizes (4.5) is given by

$$\mathbf{R} = \left(\hat{\mathbf{H}}^*\mathbf{V}^*\mathbf{V}\hat{\mathbf{H}} + \mathbf{S}^*\mathbf{S} + \overline{E}[\Delta\mathbf{H}^*\mathbf{V}^*\mathbf{V}\Delta\mathbf{H}]\right)^{-1}\hat{\mathbf{H}}^*\mathbf{V}^*\mathbf{V}\mathbf{D}. \tag{4.6}$$

The accuracy of the channel predictions are accounted for in the third term of the inverse. For details on how to calculate this term or how to derive the expression (4.6) the reader is referred to Paper II and Paper III.

In the experimental evaluation of Paper II, the sumrate performance when using (4.6) is compared to that obtained when using a zero forcing precoder

$$\mathbf{R} = \hat{\mathbf{H}}^T\left(\hat{\mathbf{H}}\hat{\mathbf{H}}^T\right)^{-1}\mathbf{D}. \tag{4.7}$$

It is important to include the diagonal target scaling matrix $\mathbf{D}$ as a right factor also of the zero forcing precoder (4.7). The sumrate performance is impaired otherwise, due to power scaling issues, in particular in fully loaded cases when $M = N$.

Before a precoder can be designed, the system needs to determine which users to serve on the individual transmission resources. On each resource, $M \leq N$ users can be grouped for JT.

From results in [4], it was clear that the performance of coherent CoMP JT in a fully loaded system $M = N$ (i.e. when the number of served users within a resource, matched that of the number of serving antennas at all base stations) was worst when all users had weak channels to the *same* base station. The conclusions that such users are spatially incompatible is intuitive, as the weak base station cannot offer much support in the transmission and hence the system is in practice overloaded. Based on this, the user grouping method proposed in Paper II, which is denoted cellular user grouping, focuses on making sure the that different users who share a resource will have *different* strongest base stations. This will make the channel matrix $\mathbf{H}$ diagonal dominant for single-antenna base station, if users are numbered appropriately. It will thus be made more easily invertible.

This aim could be attained in various ways. The here proposed method utilize an assumption that the system has cellular structure and that separate scheduling is performed for users that are assigned to each cell, performing channel aware scheduling. As a pleasant suprise, it turns out that by utilizing the results of these separate scheduling decisions, good user groups are formed spontaneously for the cluster of cells.

In this scheme, each base station schedules all users within its own cell only. The scheduling may be based on some coarse CQI of these users, e.g. a roughly quantized, resource specific, estimate of the of the channel power gain. The scheduler is assumed not to allocate a user to transmission resources in which it has an unusual weak channel. As a consequence, the scheduled users will have a reasonably strong channel to its own base station within each of the time-frequency resources it is assigned to use. Then, groups are formed for each time-frequency resource as the users that have been assigned to that resource by all the separate cellular schedulers within the cooperation cluster. As a result, users that will be jointly served on a resource will then likely have different strongest base stations, namely their own base stations [4]. The scheme has the advantage that it only needs a CQI to allocate resources and thereby forming the user groups. This lowers the feedback overhead that is needed as compared to e.g. a user grouping scheme that uses the complex valued channel gains to all users in all cells to find the optimal user groups.

[4]When one user has equally or almost equally strong channels (on average) to more than one base station, it will still have to be assigned to one of these. That base station will then be in charge of scheduling the users without any knowledge of the small scale fading of the channels between the user and the other base stations. It may therefore occur that the user has equally strong or stronger channel to a different base station for the resources that it will have been assigned. This should however not pose a problem, as the user then has two strong channel components.

The cellular user grouping may, but does not have to, utilize already existing scheduling algorithms. For example, if the objective is to maximize the sumrate, then the base station might schedule the user that has the best CQI for any given resource. If fairness amongst users is considered, then a proportional fair scheduler, which allocates the resource to the user that has the highest ratio of its CQI for the given resource relative to its average CQI for all resources, see [121], can be chosen. Another option, which will be used in the results below, is to use the score based scheduler introduced in [122].

In the evaluation in Paper II, each cell has one base station antenna and schedules at most one user in a resource. This can be generalized to multiuser MIMO transmission within cells. It need then be assumed that the mutual spatial compatibility of users that are scheduled on the same resource within a cell is taken care of by each cellular (multiuser MIMO) scheduler. These user groups are then merge into user groups for the whole cooperation cluster.

Simulation results

The proposed channel predictions, precoding design and user grouping scheme has been evaluated based on channel sounding measurements that were collected in Kista, in Sweden, by Ericsson Research in December 2008. In these, channel sounding pilots were transmitted in a 20 MHz band, at a carrier frequency of 2.66 GHz from three omnidirectional single antenna base stations located in an urban environment, see Figure 4.10.

The pilots were measured by a vehicle which was driving through the area between the base stations. The powers of the signals received from the different base stations are presented as a function of the measurement location in Figure 4.11 and as a function of time in Figure 4.12. The time series were resampled so that the true maximal vehicle velocity corresponded to a velocity of 5 km/h in the resampled time series. The quality of the measurements are very high and in the simulations included in the papers, the corresponding estimates of the OFDM channels are regarded to be the exact channels. More details on the set-up for this measurement campaign can be found in Paper II and in [123].

Based on these channel measurements, pilot measurements with different pilot SNRs were simulated for each user position plotted in Figure 4.12. These measurements were simulated through (3.2) using resource orthogonal pilots and Gaussian noise. Then, the channels were predicted for prediction horizons that correspond to distances of $\{0, 0.6, 0.13, 0.19, 0.28\}$ of the carrier wavelength in space, corresponding to system delays of $\{0, 5, 10, 15, 23\}$ ms at a user velocity of 5 km/h. The resulting NMSE was averaged over a time of 0.5 s for each user position and a total of 144 subcarriers[5].

[5]Further simulation assumptions are described in Paper II.

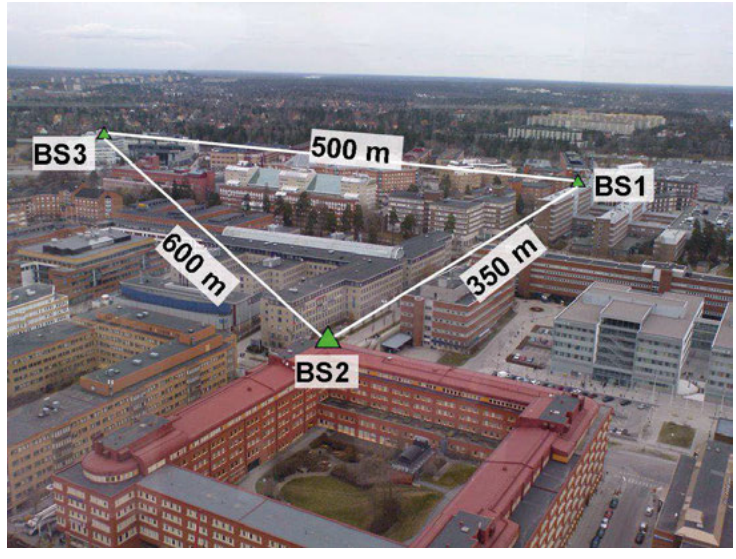

Figure 4.10. The urban environment of Kista, Stockholm, seen from above. The locations of the base stations (BS) are marked by triangles.

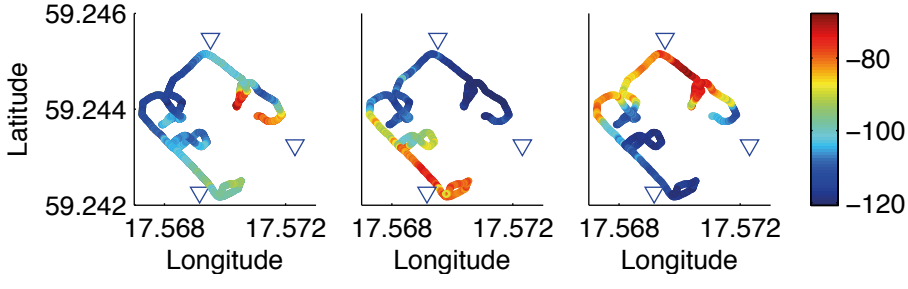

Figure 4.11. The signal powers in dBm from base station 1 (left), base station 2 (middle) and base station 3 (right). The base station locations are marked by triangles.

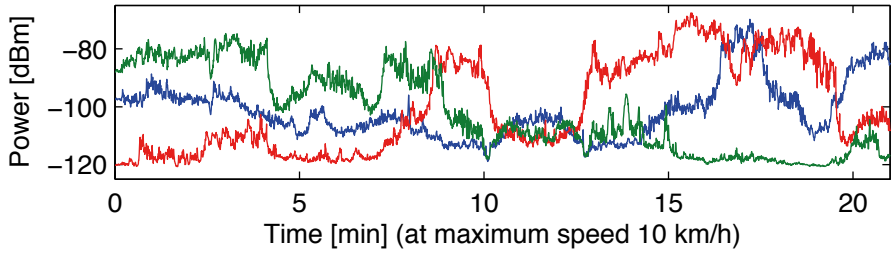

Figure 4.12. The variations of the power of the received signals that were transmitted from base station 1 (blue), base station 2 (green) and base station 3 (red). For details, see Paper II.

Figure 4.13 shows the prediction performance, in terms of average NMSE. The results are compared to the NMSE that would result from using outdated CSI, i.e. using the channel estimate $\hat{\mathbf{h}}_\tau$ as a prediction for the channel $\mathbf{h}_{\tau+m}$.

We see that while the outdated CSI very quickly decrease in accuracy, with an increasing prediction horizon, predictions through the Kalman filter provide an average NMSE ≤ -7.9 dB for prediction horizons of up to 0.28 of the carrier wavelength. As these predictions are based on real measurements and estimated channel models, that is very encouraging. It is worth noting that, as stated in Section 3.4, the measurement noise has lower impact for longer prediction horizons.

As promising the results in Figure 4.13 are, they do not show the full statistics. In Figure 4.14 the CDFs of the prediction NMSE are shown. Although most of the channels are predicted with good accuracy, there are some poor predictions among the weaker channels. For example, with the highest noise floor at -110 dB more than 30% of the channels have an NMSE of -8 dB or higher if a prediction horizon of 0.13 of the carrier wavelength is required. For the longest prediction horizon (0.28 of the carrier wavelength) this fraction increases to approximately 50%.

From Figure 4.12 it is evident that it is often the case that the signals from one or two of the base stations are much weaker that the strongest base station. Inaccurate predictions for weak channels may not be a big problem if these are very weak compared to the strong channels, as they would not contribute much to the beamforming gain.

However, it is important that these weak channels with inaccurate channel estimates do not deteriorate the quality of the total precoding solutions. This can potentially be avoided if information of the prediction accuracy is accounted for when designing the precoder.

Figure 4.15 shows the performance in terms of sumrate when the three single antenna base stations serve a group of up to three users on each of the 144 subcarriers, using joint coherent precoding. These are the sumrates that were achieved when nine users were randomly picked from the possible positions along the driving route shown in Figure 4.11 and were served jointly by the three base stations within the 144 subcarriers. The investigated scenario is a (small) fully loaded case, where the three single antenna base stations serve three users in each transmission resource, so $M = N = 3$.

First, we will focus on the solid curves of Figure 4.15. Here, the nine users were randomly assigned into group of three users, that were then randomly assigned to each subcarrier. The users were then served through coherent CoMP JT either by zero forcing precoding (circles) or though the robust linear precoder (4.6) (squares) when the precoding matrix is based on the predicted channels with a prediction horizon of 0.13 of the carrier wavelength. The criterion weight $\mathbf{S}$ in (4.5) is diagonal, and is adjusted to maximize an approximation of the sumrate. As a comparison, the sumrate that would be achieved by using zero forcing with perfect CSI is also added (diamonds).

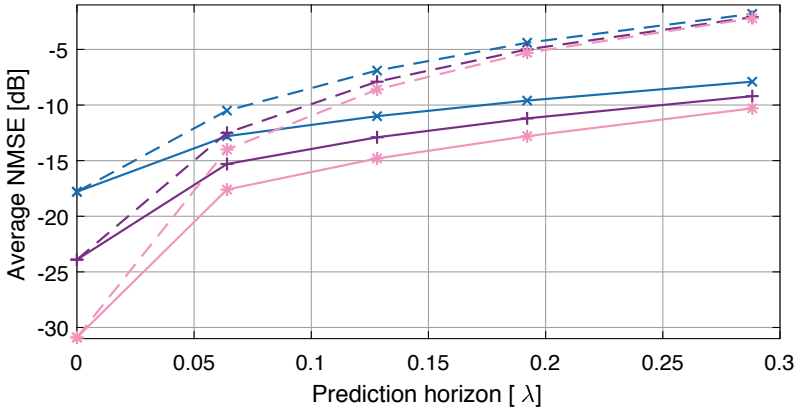

Figure 4.13. The solid lines show the prediction performance in terms of average NMSE for all base stations, user positions and subcarriers as a function of the prediction horizon scaled by the carrier wavelenght λ. Results are shown for noise levels of -110 dBm (crosses) -120 dBm (pluses) and -130 dBm (stars), see Figure 4.12. Dashed lines show the NMSE of the outdated CSI. Further details can be found in Paper II.

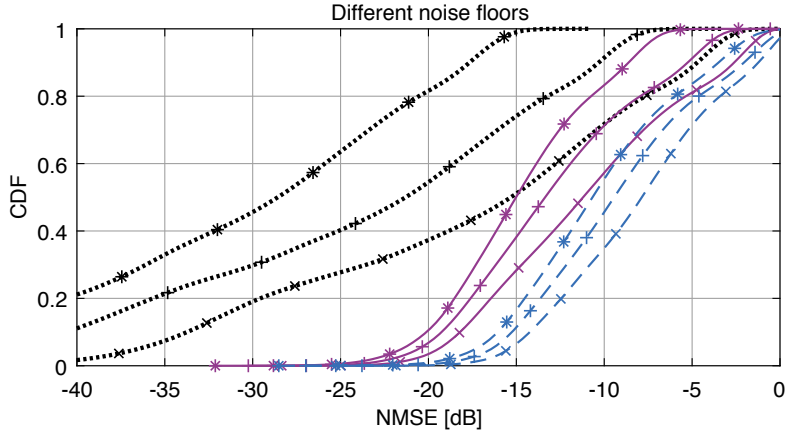

Figure 4.14. CDF of the prediction NMSE for different measurement noise levels -110 dBm (crosses) -120 dBm (pluses) and -130 dBm (stars). Prediction horizons of 0 carrier wavelengths (black dotted lines), 0.13 carrier wavelengths (purple solid lines) and 0.28 carrier wavelengths (blue dashed lines). For further details, see Paper II.

We can see, by comparing zero forcing with robust linear precoding, that knowledge about the prediction accuracy is important in this situation, especially when we compare the low percentiles in the CDFs at high noise floor (-110 dBm). Here, the sumrate is increased significantly. The gap is lowered as the noise level decreases. However at no point does the robust linear precoder provide a worse sumrate than zero forcing.

Now consider instead the dotted curves, representing results for the nine users that are grouped by cellular scheduling with score based schedulers. That is, for each simulation, the users that have the *same* strongest base station are assumed to belong to the same cell and are therefore scheduled on *different* subcarriers using the score based scheduler introduced in [122]. Users scheduled on the same resource are then served jointly by the base stations.

This simple user grouping scheme has the effect that the problem with bad user grouping becomes insignificant. Comparing the dotted and the solid curves, the large impact of using a reasonable user grouping scheme becomes evident. First, and most striking, there is a large improvement in sumrate performance. Second, the impact of different precoding schemes is altered. At the lowest noise floor (-130 dBm), zero forcing now even outperforms robust linear precoding. At first, this may seem suspicious, however it is a case of being overly confident in the information we have. Recall from Section 3.4 that the channel model is estimated from imperfect training data. That means that the information we have in regards to the accuracy of the channel prediction is slightly off. For most theoretical scenarios, this is not an issue as the estimated accuracy is not far from that of the actual accuracy, see e.g. Chapter 8 of [1] where this was investigated for the channel data set used here. However, as the system is pushed to the bounds, as in this case with very low noise floor and very good user groups, the errors made in estimating the prediction accuracy will cause the robust linear precoder to actually perform somewhat worse than zero forcing.

In addition to the results presented here, Paper II provides simulation results where the proposed cellular user grouping method presented here is compared to a high complexity, greedy user grouping scheme, which iteratively adds a users to the group within each resource. The simulations, which are based on block fading channels, show very little performance loss with cellular user grouping as compared to the greedy user grouping scheme.

4.2.3 Handling backhaul limitations

Depending on the infrastructure, the requirements on the backhaul might not always be manageable, and hence we need a method for handling constraints on how much information that can be sent over the backhaul links.

Limitations in how much CSI can be fed to the CU can be handled by setting the unavailable channel estimates to zero and adjusting the covariance matrices

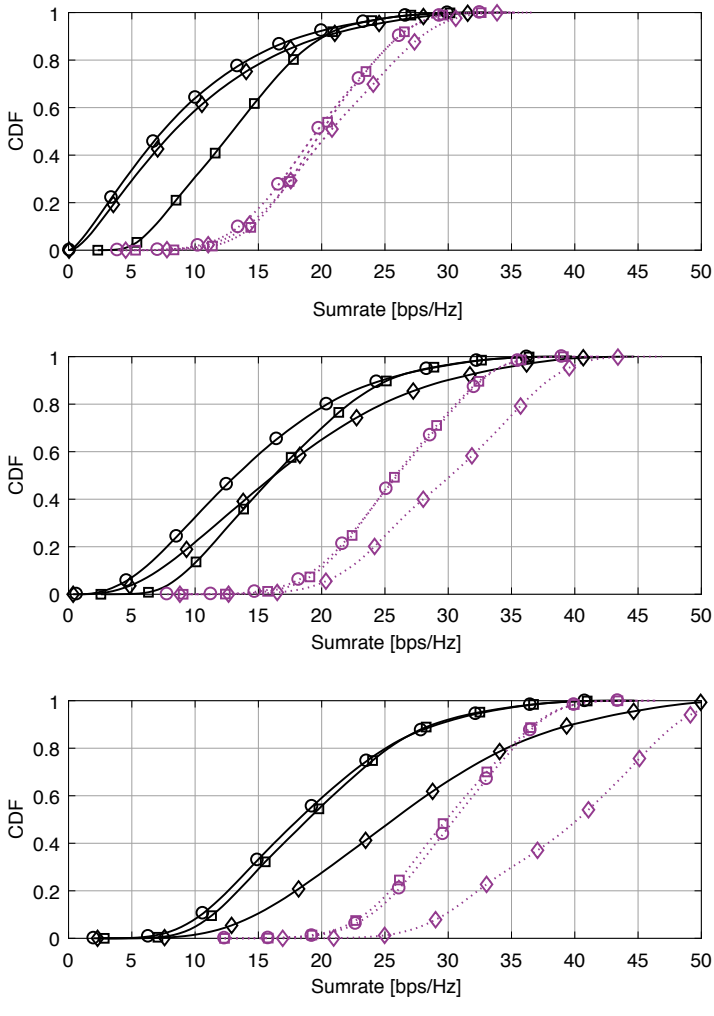

Figure 4.15. CDFs of the sumrate for all user groups when nine users are randomly dropped along the route shown in Figure 4.11 and grouped over 144 subcarrier with up three users to be jointly served per subcarrier. Results are shown when the precoding matrix is obtained through zero forcing precoding (circles) or through the robust linear precoder (4.6) (squares). Users are either grouped randomly and then allocated resources randomly, by so called round robin scheduling, (solid lines), or by cellular user grouping, when each of three cells allocates its users using a score based scheduler (dotted lines). Results for noise floors of -110 dBm (top), -120 dBm (middle) and -130 dBm (bottom). Results with perfect interference suppression by zero forcing with perfect CSI (diamonds) is added for comparison. For further details, see Paper II.

of the error accordingly, see Paper III for details. As the error covariance matrices only need to be sent on a slow time scale, e.g. every half second, these place very low demands on the backhaul capacity.

However, we may also need to limit the number of non-zero elements in the precoding matrices that are distributed to the base stations.

Furthermore, assume that the sharing of payload data between sites is limited so that only some of the payload can be shared. Then, some base stations will not have access to all the symbols of the data vector $\mathbf{s}$. This can be mathematically modelled the same way as when a limited number of precoding weights can be fed back to the base stations from the CU, i.e. by setting the appropriate elements of the precoding matrix $\mathbf{R}$ to zero. That is, if no data is available at base station antenna n for user m, then element $\{n,m\}$ of the precoding matrix, $r_{n,m}$, should be set to zero.

In Paper III, this is done by extending the criterion (4.5) to

$$J = \bar{E}\left[E\left\|V\varepsilon\right\|^2 + E\left\|Su\right\|^2 + E\sum_{n=1}^{N}\sum_{m=1}^{M}\frac{1}{c^2}\left|w_{nm}r_{nm}s_m\right|^2\right], \tag{4.8}$$

where s_m is the message symbol for the m'th user and the scalar w_{nm} is a penalty weight that is set high if the element r_{nm} of the precoder matrix $\mathbf{R}$ needs to be zero and is set to zero otherwise.

The solution which minimizes this criterion is found by first calculating the precoder that minimizes the original criterion, which is given by

$$\mathbf{R}_0 = (\beta^*\beta)^{-1}\hat{\mathbf{H}}_{CU}^*\mathbf{V}^*\mathbf{V}\mathbf{D}, \tag{4.9}$$

where $\hat{\mathbf{H}}_{CU}$ is the available estimate of the channel matrix at the CU and

$$\beta^*\beta = \hat{\mathbf{H}}_{CU}^*\mathbf{V}^*\mathbf{V}\hat{\mathbf{H}}_{CU} + \mathbf{S}^*\mathbf{S} + \overline{E}[\Delta\mathbf{H}_{CU}^*\mathbf{V}^*\mathbf{V}\Delta\mathbf{H}_{CU}]. \tag{4.10}$$

Then the precoder with (approximately) enforced zeros is given by solving a linear matrix equation

$$\mathbf{R} = \mathbf{R}_0 - (\beta^*\beta)^{-1}(\mathbf{W}\odot\mathbf{W}\odot\mathbf{R}). \tag{4.11}$$

Here $\mathbf{W}$ is the matrix given by the weights w_{nm} and $\odot$ denotes elementwise multiplication.

Results in Paper III are based on a simple simulation environment with three sites, each with three sectored base stations with two antennas each and nine randomly distributed single antenna users with at most one allocated user per cell and transmission resource. A large number of sets of user positions is simulated. The channels are modelled as block fading and channel errors are modelled as complexed-valued zero mean Gaussian variables with a variance which is 20 dB below the channel variance. No shadow fading is considered. The results in Paper III show that there is a gain in the average user rate of up to 13% when using the proposed precoder with the extended MSE criterion,

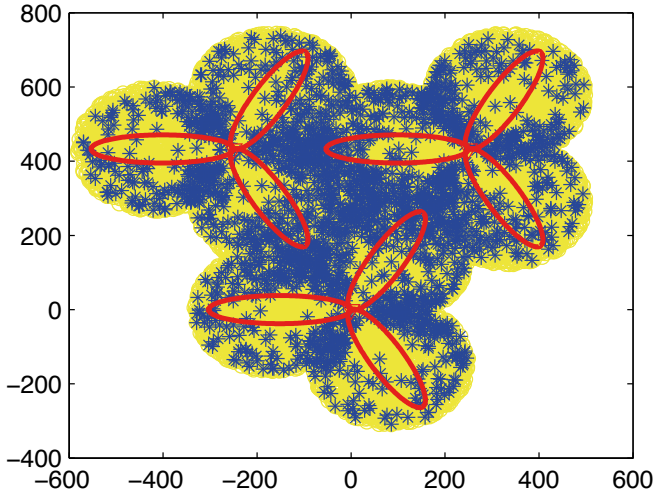

Figure 4.16. The user positions that benefit by more than 50% in terms of sumrate from using the precoder design proposed in Paper III, are shown as blue stars. The precoder constraint imposed here is that each user should only be served by its four strongest base stations. The base station antenna directions are provided by the red ellipses and all other user positions are marked as yellow circles. For more details, see Paper III.

as compared with inserting zeros as a last step, i.e. by simply changing the appropriate elements of the precoding weights to zero after having calculated the precoder.

However, as CoMP is especially important for cell edge users, it is of great interest to limit the effect of backhaul constraint for them. The proposed precoder does exactly that. For example, in a scenario where each user is served only by its four strongest base stations, in a cluster with nine base stations, then 13% of the users obtained a capacity gain of more than 50% when using the proposed precoder with the extended MSE criterion, as compared with inserting zeros as a last step, i.e. by simply changing the appropriate elements of the precoding weights to zero after having calculated the precoder. These users, which are plotted in Figure 4.16, are mainly cell edge users.

4.3 Channel smoothing for TDD systems with predictor antennas

In a TDD system, channel reciprocity can be used to estimate the downlink channel based on resource orthogonal uplink pilots from the scheduled users. As the number of user may be considerably less than the number of transmit

antennas for massive MIMO systems, the pilot overhead may be considerably smaller than in an FDD system even if resource orthogonal pilots are used.

With wireless access playing a more and more important role in peoples daily life, problems may occur when many users are a gathered in the same small location, creating so called hot-spots. A typical hot spot scenario occurs within the sector of public transportation such as trams or buses where people use their mobile wireless communication equipment both for work and for entertainment purposes [124].

While linear prediction works well for pedestrian users, it may not be sufficient for high mobility users when the delays between the pilots and the data transmission are of significant duration. An example is a TDD system where the user equipment would transmit known pilots during an uplink frame, that are used for channel estimation on the network side. Assuming channel reciprocity, these channel estimates are then used to calculate predictions of the channels to the user equipment antennas during a subsequent downlink subframe. The required prediction horizon for the downlink channels could then be on the order of the magnitude of the carrier wavelength in space, for vehicular users.

An approach to increase the prediction horizon might be to have a database of pre-recorded coordinate specific CSI, e.g. located at the base stations [125]. The users can then feed back an estimate of their locations, e.g. based on Global Positioning System (GPS) information, to the base stations. As this method would not really be predicting the fading in the conventional sense, it is not bound by the limitations discussed in Section 3.4. However, such a model would require collection of a vast amount of data on centimetre scale in space. This data might in turn need to be updated continuously, due to e.g. seasonal changes in the environment. It is also unclear how such a scheme might be affected by e.g. bypassing vehicles that alter the standing wave pattern. It is therefore of interest to investigate more easily realizable long-range channel prediction methods for vehicular users.

4.3.1 History of the predictor antenna concept

The limit of prediction horizon can be circumvented by the predictor antenna concept which is illustrated in Figure 1.4. It uses an extra antenna, predictor antenna, placed in front of one or several main antennas, e.g. on the roof of the vehicle, to scout the channel that will later be encountered by the main antennas. This concept was originally proposed by Sternad et al. in [27], where the potential of predictor antennas were illustrated based on a small measurement campaign with one LOS and one NLOS scenario, using two dipole antennas on the roof of the vehicle.

A second measurement campaign with monopole antennas and a large metal sheet under the antennas showed improved results [43, 126, 127]. Further ex-

perimental studies, conducted by Björsell, Sternad and Grieger have shown a prediction NMSE of about -10 dB for vehicular velocities for all measured predictions horizons up to three times the carrier wavelenght in space, ten times longer than the limit for channel extrapolation [128, 129]. The method has recently been applied to 64-antenna massive MIMO downlinks, and was shown to provide close to ideal maximum ratio transmit beamforming gains to vehicles over NLOS channels [130].

To produce accurate predictions, the concept requires channel estimates for the predictor antenna from positions close to where the main antenna will transmit or receive signals. For a FDD system, this requires dense enough downlink channel estimates, which was the case in [128, 129, 130], where OFDM pilots, evenly distributed in time and frequency, were used.

Similar antenna systems as in [27] with two in-line antennas on a vehicle roof, were investigated in the 90's by Vaughan and coworkers [131, 132]. The aim was there not prediction, but to model the radio environment. These experimental results showed that the maximal cross correlation between the signal envelope at the rearward antenna and the delayed signal envelope at the forward antenna was reduced when the antennas were separated by 0.2-0.4 times the carrier wavelength, due to antenna coupling. In [132], the effects of the antenna coupling were compensated for a to a large extent. Similar results were found in [43, 126, 127] and the antenna coupling was sucessfully compensated for in [127].

4.3.2 Kalman smoothing for TDD downlink estimates

Even when using predictor antennas, the distance between the antennas may not be perfectly aligned with the required prediction horizon (as this will depend on the users speed). For a TDD system, the uplink/downlink ratio to the TDD frame might be adjusted so that the downlink transmission of the main antenna occurs close to a position where the predictor antenna already has measured the channel, as proposed in [28] and evaluated in [133]. However, such a scheme would require individually adaptable uplink/downlink ratios based on the velocity of each user which is problematic from a system perspective. Instead, as suggested in [133], interpolation can be used for any given uplink/downlink ratio to generate channel estimates for the gaps in the uplink pilot sequence.

The results of Paper V, which are highlighted in this section utilize Kalman smoothing to interpolate the channels during TDD downlink frames, by using received uplink pilot signals from the predictor antenna in uplink frames. The interpolations use a two-filter approach, using two state space models.

The Kalman smoother is the MMSE-optimal linear interpolator of noisy data, for known second order statistics of signal and noise [10]. It has been studied in applications such as the compensation for packet loss in wireless

sensor network system [134] and for channel equalization on the receiver side [135], in both cases where partial observation losses occur.

To gain understanding of the concept, let us consider a predictor antenna system with two antennas, spaced by a fixed distance and placed in a straight line in the direction of travel on a vehicle, as was illustrated in Figure 1.4. We define a time interval indexed by τ, which shall here be denoted time slot and which represents the time between potentially pilot bearing OFDM symbols. (Between these OFDM symbols separated by τ, data bearing OFDM symbols are in general be transmitted.)

Assume that required prediction horizon is m time slots and assume that the vehicle is travelling with a velocity such that it takes $m+\mu$ time slots for the main antenna to reach the position of the predictor antenna. As the main antenna will experience approximately the same channel as the predictor antenna, with a time lag of $m+\mu$ time slots, the prediction of the main antenna's channel at time $\tau+m$ can be given as a smoothed estimate of the predictor antenna's channel at time $\tau-\mu$. We can therefore use up to μ later channel estimates to improve the estimate of the channels from the predictor antenna at time τ, if measurements are available at these later time slots.

An example of this is illustrated in Figure 4.17, where a channel from one base station antenna in a TDD system is to be estimated. Here, pilots are transmitted in the uplink for three consecutive pilot slots, indexed by τ, followed by a downlink with three consecutive slots of equal length to the uplink slots but without pilots. In this example it is assumed that $m=\mu=5$ time slots and that the vehicle is travelling at a speed where five time slots in time corresponds to one carrier wavelength in space. Let us focus on time index $\tau=9$, when the aim is to predict the channel for the main antenna at time index $\tau=14$, which is given by position four, as shown in Figure 4.17. At this point in time, the system has access to CSI based on the predictor antenna pilots at positions 1-3 and 7-9, but not for position four. There is no measurement for that position, so interpolation will be needed.

The relationship between the NMSE of the smoothed estimate of the predictor antenna channel NMSE_p and the NMSE of the prediction for the main antenna channel NMSE_m is given by equation (14) in [128]:

$$\mathrm{NMSE}_m = 1 - \frac{|b|^2}{1+\mathrm{NMSE}_p}, \tag{4.12}$$

where b is the maximum normalized cross-correlation between the channel of the main antenna and the delayed channel of the predictor antenna. This parameter is $=1$ in the ideal case, and it determines the ultimate performance for error free measurements of the channel of the predictor antenna. The estimation error to channel power ratio NMSE_p also influences the performance, by (4.12). It is therefore of importance to obtain good smoothed estimates of the channel for the prediction antenna.

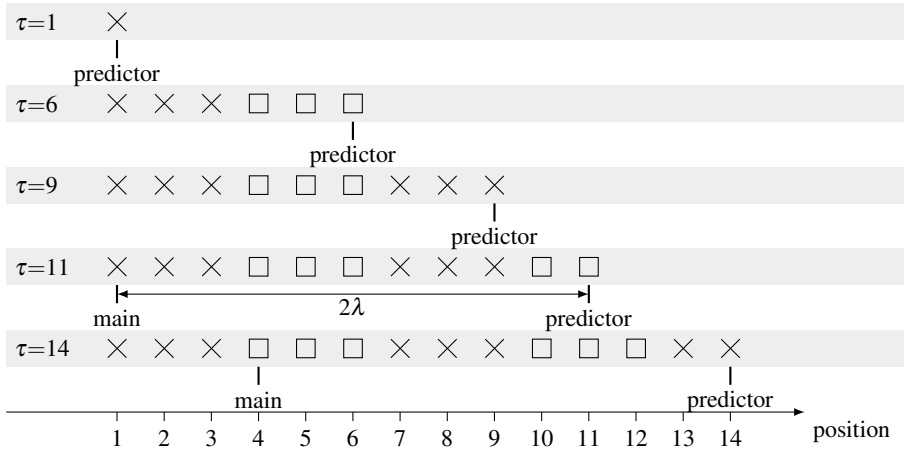

Figure 4.17. A TDD system where pilots are transmitted in the uplink for three consecutive pilot slots followed by a downlink subframe with three consecutive slots without pilots. The figure shows the positions of the predictor antenna and of the main antenna at different slot times τ. Crosses mark positions for which the predictor antenna has transmitted pilots and thus made channel estimates availible.

In the following, the focus will be on the NMSE of the smoothed estimates of the channel from the predictor antenna, when using the two filter solution described in Paper V and in Section 3.2.1.

4.3.3 Important results and conclusions

To evaluate the scheme we use measurements collected while driving at a speed of 25-39 km/h in downtown Dresden. Channel sounding pilots were transmitted every 0.5 ms from four antennas mounted on the roof of the measurement van at a carrier frequency of 2.53 GHz (λ = 119 mm) and were received at a base station. The set-up is shown in Figures 4.18 and 4.19.

For the results presented here, three measurement sets obtained at high SNR were used. These represent three different scenarios; A LOS scenario, a NLOS scenario with a Doppler spectrum similar to that of Rayleigh fading and a NLOS scenario with a relatively flat Doppler spectrum, see Figure 4.20. The measurements are selected from a larger set, which is described in greater detail in [128].

The original channel measurements were filtered for noise and were subsampled such that they would correspond to the channels collected when pilots are transmitted every 1 ms at a vehicle velocity of 75-83 km/h, see Paper V for details.

As stated above, the focus here is on the estimation of the predictor antenna channel, so for the purpose of these investigations, all four antennas in the measurement campaign will be considered predictor antennas. An extension

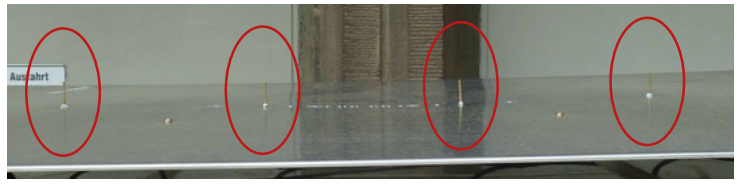

Figure 4.18. Antenna set-up for the measurements collected in downtown Dresden. Four monopole antennas, spaced by two times the carrier wavelength were mounted over a metal sheet to ensure that they experience similar fading.

Figure 4.19. Measurement set-up for the measurements collected in downtown Dresden. The metal sheet with the four antennas was placed on top of a measurement van.

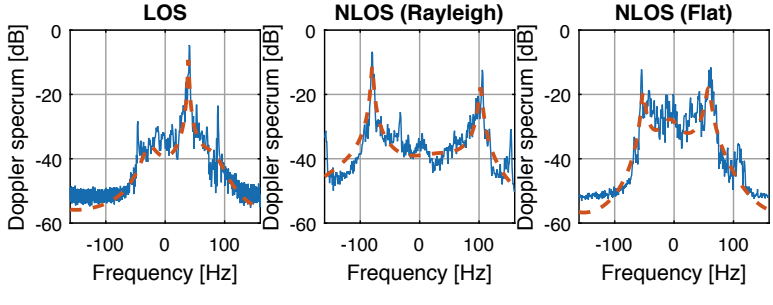

Figure 4.20. Doppler spectra of the three measurements used in the evaluations are provided as a solid lines, while the Doppler spectra of the estimated AR models are represented by the dashed line. The Doppler spectra shown here are those found from the measurements of the first antenna.

of these investigations, where the three first antennas are used as predictor antennas, each for the antenna that is placed behind it, and the cross-correlations between each pair of main/predictor antennas are used to translate the predictor antennas' smoothed channel estimates into a channel predictions for the main antennas are left as future work.

To relate the results presented here to what could be achieved by the full system, if $|b|$ is close to $|b| = 1$, then by (4.12) $\text{NMSE}_m \approx \text{NMSE}_p$ for $\text{NMSE}_p \leq -8dB$.

Based on these processed measurements, which in the simulations are considered accurate representations of the radio channels, measurement signals were simulated by (3.2) with added white noise simulating pilot SNRs of -5, 5 and 15 dB. To model a TDD system with frame structure as in Figure 4.17, the pilot matrix Φ_τ of (3.2) was set to a unit matrix for $\tau = \{1,2,3\} + 3m$ and an all zero matrix for $\tau = \{4,5,6\} + 3m$ with m being an integer.

The small scale fading of each channel was modelled by fourth order AR models. Figure 4.20 shows how well the models fit the Doppler spectra. These where then used in the Kalman smoothers to find the channel estimate $\hat{\mathbf{h}}_{\tau|\tau+5}$, assuming a required prediction horizon of one carrier wavelength, an antenna spacing of twice the carrier wavelength and a pilot sampling interval of 1 ms.

In the scenario described in Figure 4.17 above, the channel estimate of interest (used as predictor for the main antenna channels) is at the position where the predictor antenna was 5 time steps before it reached its current position. We therefore evaluate the Kalman smoothing estimate $\hat{\mathbf{h}}_{\tau|\tau+5}$ for the predictor antenna. The NMSE will depend on the pilot SNR and on the location of τ inside or outside of the downlink subframes. For $\tau = \{1,2,3\}$ it is within an uplink subframe, where pilots are available. For $\tau = \{4,5,6\}$ we target a point that was passed during a downlink subframe.

Results are shown in Figure 4.21. Here, the smoothed estimate is compared to a Kalman *filter* estimate $\hat{\mathbf{h}}_f(\tau|\tau)$. Note that the filter estimate will constitute a predictor from past estimates for $\tau = \{4,5,6\}$. Furthermore, results are included for a scenario where estimates are obtained by using the measurements at $\tau = \{1,2,3\} + 3m$ and performing smoothing cubic interpolation through the matlab function `csaps`, see Paper V for further details.

The NMSE shown in Figure 4.21 is averaged over 88 subcarriers and four transmit antennas. As a benchmark for estimation performance, we can use an NMSE of ≤ -8 dB, which will result in good prediction performance for a prediction horizon of one wavelength assuming that $|b|$ is close to $|b| = 1$ in (4.12).

For this particular scenario, our findings can be summarised as follows

- Smoothing cubic interpolation is insufficient for estimating all downlink channels in any of the scenarios, whereas Kalman smoothing ensures good estimation performance for both the LOS scenario and the NLOS with Rayleigh fading at an SNR of 5 dB and above.

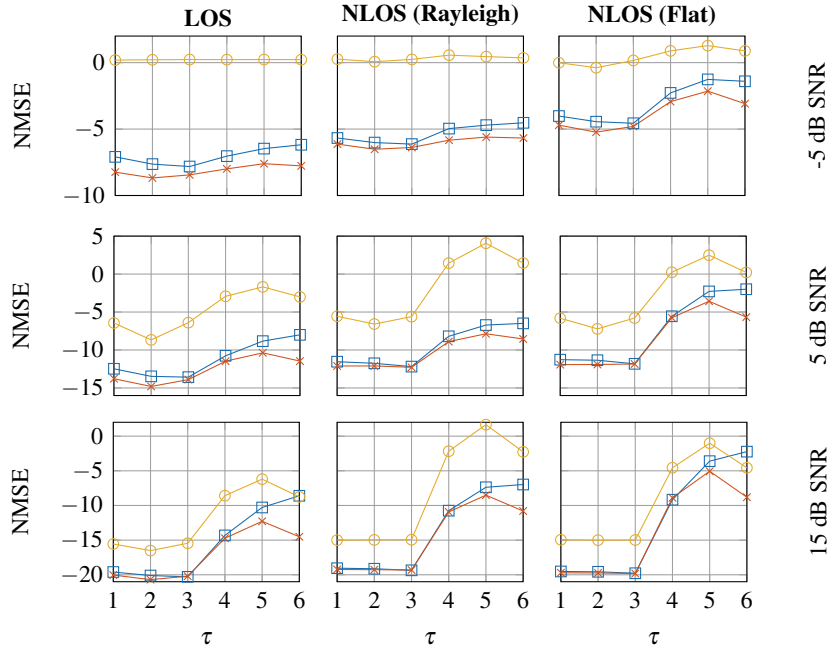

Figure 4.21. The average NMSE for $\tau = 1,\ldots,6+6m$ where m is an integer, corresponding to 1 ms. Results are shown for Kalman filtering (squares), Kalman smoothing (crosses) and interpolation through smoothing cubic interpolation (circles). Results are for the different types of channels of Figure 4.20 (figure columns) and for different SNR levels (figure rows).

- An added bonus from Kalman smoothing is the noise reduction at low SNR. For the LOS scenario this ensures an NMSE below -8 dB for time slots $\tau = \{1,2,3\}$ at an SNR of -5 dB, which the filter/predictor does not achieve.
- When the temporal correlation is lower, as in the case of the flat Doppler spectrum, it is harder to achieve the here targeted NMSE, even with smoothing.
- The performance difference between smoothing and filtering is largest for τ within downlink subframes $\tau \in \{4,5,6\}$. Here, the filter/predictor $\hat{\mathbf{h}}_{\tau|\tau}$ is forced to perform a prediction based on previous pilot bearing symbols, while the smoother $\hat{\mathbf{h}}_{\tau|\tau+5}$ also utilizes measurements from the next uplink subframe.

To summarize, the results indicate that downlink subframes should be of a duration that corresponds to no longer than 0.6-0.7 of the carrier wavelength in space at the highest user velocity that the system is required to support. These results are in line with those of Section 3.4 and Section 4.2, where predictions were shown to provide NMSE of -8 dB or lower for horizons up to 0.3 of the carrier wavelength. Since the smooting of the downlink channels in

essence predicts from two directions in time and weighs these two predictions together, it is reasonable that this should apply for a distance which is close to or slightly better than twice of the prediction limit, when predicting only from one direction.

The results here focus on when Kalman smoothing is used to estimate the downlink channels based on uplink pilots in a TDD system. However, the smoother could also be used for other applications. Such an example could be to estimate the channels in between pilot samples in a system where pilots are sparse in space (either because the user is travelling very fast or because the pilots are transmitted very sparsely in time).

5. Conclusions

The Kalman filter has been the foundation of the channel estimations in this thesis. By modelling the small scale fading of radio channels by low order AR models, Kalman filters can be used to obtain filter estimates, predictions and/or smoothed estimates of these radio channels at a reasonable complexity, while taking into account correlations in time, frequency and across different physical radio channels.

The pilot structure affects the estimation performance. While the best performance is achieved from resource orthogonal pilots, non orthogonal pilots may sometimes be necessary to keep pilot overhead within reasonable levels, e.g. when estimating channel downlinks in a massive MIMO FDD system. In such a case, it is useful to design the system so that the limited amount of pilots will be sufficient to get acceptable estimation performance.

One such way is by introducing a fixed grid of beams. By directing the pilot energy into different directions, the number of channel components that need to be estimated, as seen from the point of a specific user, will then be reduced. By using pilot codes, over the available pilot resources, that are linearly independent for any set of beams lower than or equal to the number of available pilot resources, any users can then estimate at least that many channels, provided that the channels that are not to be estimated are weak in comparison.

To improve estimates further, Pilot codes can be cycled over time, which enhances the estimates when users are able to utilize temporal correlation through Kalman filtering.

Simulation results showed that when using a fixed grid of beams in combination with the proposed cycling pilot codes, a reduced order Kalman filter (which only tracks the strongest channels) provided an estimation NMSE for the channels to the beams in a CoMP cluster with massive MIMO antennas that was 5 dB above (worse) than if orthogonal pilots were used. The gain from using the temporal correlation through Kalman filtering (compared to an LLMSE estimate based on the current measurement only) was 5 dB in terms of NMSE, in the investigated example. This gain will in any particular case be dependent on the user mobility and the pilot rate.

Previouse results [1, 14] suggest that the prediction horizon that can be achieved by the Kalman filter is limited to a few tenths of the carrier wavelength in space. Depending on the carrier frequency and the user mobility, this translates to different delays in time. For long time delays (of tens of ms) and carrier frequencies in the GHz range, this limit means that channel predictions for vehicular users mobilities would be very difficult to achieve.

Generally channels with a peaky Doppler spectrum (such as a LOS channel, a Rayleigh fading channel or a channel with a few strong NLOS components) are easier to predict than one with more flat Doppler spectra. The main limitations to the predictability arise from the quality of the training data that is available. Both noisy training data and a time windowing of the training data greatly limit the prediction performance of an estimated channel model. As these two impairments cannot be avoided - noise will always be a factor even with very advanced receiver filters and the amount of available training data is limited to the time that the channel remains semi static, which in turn depends on the user mobility - it is impossible to predict channels infinitely into the future. In fact, both results based on simulated channels in Section 3.4 and those based on channel measurements in Section 4.2 support the limited prediction horizon reported in the works of [1, 14].

The second factor that limits the predictability of a radio channel is the model order with which the small scale fading is modelled. If the autocorrelation function is perfectly known, then a higher model order provides a better estimate. However with limited training data, even a lower model order will often be a better choice, suggesting that for high mobility users, a low model order is a more suitable choice, not only from the perspective of keeping computational complexity low.

The measurement noise of the pilot signals generate a third order effect for long term predictability. Noise on the received pilots is most important for the filter estimate and the short range prediction horizons.

Kalman based channel predictions for slow users (of approximately 10 km/h) were used in Section 4.2 to evaluate performance in CoMP systems with long system delays. Here, not only the channel prediction, but also information regarding the accuracy of these were included in a proposed robust linear precoder design. Measurement based simulations showed that if users were randomly grouped and then served by a CoMP cluster of three single antenna base stations by coherent JT, then it is of great importance that the information about the prediction accuracy is included in the precoder.

However, when the much superior cellular user grouping scheme, proposed in Section 4.2, is assumed, similar CoMP gains, in terms of sumrate, can be achieved also with a zero forcing precoder.

The proposed robust linear precoder has a second advantage. It is easy to adjust the the minimization criterion of the precoder to handle backhaul constraint. Simulation results suggests that this is especially important so that cell edge users do not suffer unnecessarily from backhaul constraints.

When the prediction horizons required in downlink transmission to vehicles are longer than what can be achieved with Kalman filters, a predictor antenna can be useful. This is an antenna that if placed on the roof of the vehicle in front of the main antenna. The predictor antenna can then scout the channel, efficiently predicting the channel that the main antenna will receive later.

Depending of the distance between the antennas, the predictor antenna may be able to provide measurements to enable a smoothed channel estimate of the main antenna's future channel. This is especially important is for some reason no pilots are transmitted during a longer time period. An example of such a scenario is in a TDD system where downlink channels are based on uplink pilot measurements. Then the channels within the downlink frames much either be extrapolated based on the uplink pilots or, if "future" estimates are made available as a result of a predictor antenna, interpolated.

Using Kalman smoothing to interpolate these, provides good channel estimates for downlink frames of a duration that corresponds to approximately 0.6 of the carrier wavelength in space.

References

[1] D. Aronsson, *Channel estimation and prediction for MIMO OFDM systems - Key design and performance aspects of Kalman-based algorithms*. PhD thesis, Uppsala University, Mar. 2011. http://www.signal.uu.se/Publications/ptheses.html.

[2] EU FP7-ITC-2009, ARTIST4G project report D1.4, "Interference avoidance techniques and system design," 2012.

[3] J. Li, A. Papadogiannis, R. Apelfröjd, T. Svensson, and M. Sternad, "Performance analysis of coordinated multi-point transmission schemes with imperfect CSI," in *Proc. of IEEE PIMRC 2012*, (Sydney, Australia), 2012.

[4] R. Apelfröjd, M. Sternad, and D. Aronsson, "Measurement-based evaluation of robust linear precoding for downlink CoMP," in *Proc. of IEEE ICC 2012*, (Ottawa, Canada), pp. 254–259, June 2012.

[5] W. Zirwas, M. B. Amin, and M. Sternad, "Coded CSI reference signals for 5G - exploiting sparsity of massive MIMO radio channels," in *20th International ITG Works. on Smart Antennas (WSA 2016)*, (Munich), Mar. 2016.

[6] Fraunhofer Heinrich Hertz Institute, "Quasi deterministic radio channel generator user manual and documentation," 2016.

[7] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[8] E. Dahlman, S. Parkvall, and J. Sköld, *4G: LTE-Advanced Pro and The Road to 5G*. Academic Press, 2016.

[9] Qualcomm AB, "5G radio access," *Qualcomm White Papers*, Dec. 2016.

[10] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Prentice Hall, 2000.

[11] M. Sternad, S. Falahati, T. Svensson, and D. Aronsson, "Adaptive TDMA/OFDMA for wide area coverage and vehicular velocities," in *Proc. IST Mobile and Vehicular*, (Dresden, Germany), June 2005.

[12] R. Fritzsche, P. Rost, and G. Fettweis, "OFDM channel estimation by singular value decomposition," *IEEE Trans. Wireless Commun.*, vol. 14, pp. 4417 – 4427, 2015.

[13] N. Ravindran and N. Jindal, "Multi-user diversity vs. accurate channel state information in MIMO downlink channels," *IEEE Trans. Wireless Commun.*, vol. 11, pp. 3037–3046, 2012.

[14] T. Ekman, *Predictions of Mobile Radio Channels*. PhD thesis, Uppsala University, Oct. 2002. http://www.signal.uu.se/Publications/ptheses.html.

[15] P. Papoulis, "A note on the predictability of band-limited processes," *Proceedings of the IEEE*, vol. 73, pp. 1332 – 1333, 1985.

[16] F. Marvasti, "Comments on "A note on the predictability of band-limited processes"," *Proceedings of the IEEE*, vol. 74, p. 1596, 1986.

[17] K. E. Baddour and N. C. Beaulieu, "Autoregressive modeling for fading channel simulation," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 1650 – 1662, 2005.

[18] R. H. Clarke, "A statistical theory of mobile-radio reception," *The Bell System Technical Journal*, vol. 47, pp. 957 – 1000, 1968.
[19] T. Ekman, "Analysis of the LS estimation error on a Rayleigh fading channel," in *Proc. of IEEE VTC Spring 2001*, (Rhodes, Greece), May 2001.
[20] S. Jaekel, L. Raschkowski, K. Borner, and L. Thiele, "Quadriga: A 3-D multi-cell channel model with time evolution for enabling virtual field trials," *IEEE Trans. on Antennas and Propagations*, vol. 62, pp. 3242–3256, 2014.
[21] A. Papadogiannis, E. Hardouin, and D. Gesbert, "Decentralising multicell cooperative processing: A novel robust framework," *EURASIP Jour. on Wirel. Com. and Netw.*, 2009.
[22] J. Li, *Downlink resource allocation in cooperative wireless networks*. Ph.d. thesis, Chalmers University of Technology, Dec. 2014.
[23] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Trans. on Inf. Theory*, vol. 50, pp. 1893–1909, 2004.
[24] W. Ajib and D. Haccoun, "An overview of scheduling algorithms in mimo-based fourth-generation wireless systems," *IEEE Netw.*, vol. 19, pp. 43–48, 2005.
[25] T. Yoo and A. Goldsmith, "Optimality of zero-forcing beamforming with multiuser diversity," in *IEEE ICC 2005*, (Seoul, South Korea), May 2005.
[26] M. Joham, W. Utschick, and J. A. Nossek, "Linear transmit processing in MIMO communications systems," *IEEE Trans. Signal Process.*, vol. 53, pp. 2700–2712, 2005.
[27] M. Sternad, M. Grieger, R. Apelfröjd, T. Svensson, D. Aronsson, and A. B. Martinez, "Using predictor antennas for long-range prediction of fast fading for moving relays," in *Proc. of IEEE WCNC 2012*, (Paris, France), Apr. 2012.
[28] D.-T. Phan-Huy and M. Helard, "Large MISO beamforming for high speed vehicles using separate receive and training antennas," in *Proc. of IEEE ISCWS 13*, (Ilmenau, Germany), Aug. 2013.
[29] R. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME (J. Basic Engineering)*, vol. 82, pp. 35–45, 1960.
[30] O. Edfors, M. Sandell, J.-J. van de Beek, S. K. Wilson, and P. O. Börjesson, "OFDM channel estimation by singular value decomposition," *IEEE Trans. Commun.*, vol. 46, pp. 931 – 939, 1998.
[31] P. Hoeher, S. Kaiser, and P. Robertson, "Two-dimensional pilot-symbol-aided channel estimation by wiener filtering," in *Proc. of IEEE ICASSP97*, (Munich, Germany), Apr. 1997.
[32] L. Lindbom, "Simplified Kalman estimation of fading mobile radio channels: high performance at LMS computational load," in *Proc. of IEEE ICASSP 93*, (Pacific Grove, USA), Apr. 1993.
[33] C. Komnimakis, C. Fragouli, A. Sayed, and R. Wesel, "Multi-input multi-output fading channel tracking and equalization using kalman estimation," *IEEE Trans. Signal Process.*, vol. 50, pp. 1065–1076, 2002.
[34] M. Sternad, T. Svensson, T. Ottosson, A. Ahlen, A. Svensson, and A. Brunström, "Towards system beyond 3G based on adaptive OFDMA transmission," *Proc. of IEEE*, vol. 95, pp. 2432 – 2455, 2007.
[35] D. Aronsson and M. Sternad, "Kalman predictor design for frequency-adaptive

scheduling of FDD OFDM uplinks," in *Proc. of IEEE PIMRC 07*, (Athens, Greece), Sept. 2007.
[36] S. M. Kay, *Fundamentals of Statistical Signal Processing. Volume 1: Estimation Theory*. Prentice Hall, 1993.
[37] F. Gustafsson, L. Ljung, and M. Millnert, *Signal Processing*. Lund, Sweden: Studentlitteratur AB, 2010.
[38] K. Manolakis, S. Jaeckel, V. Jungnickel, and V. Braun, "Channel prediction by doppler-delay analysis and benefits for base station cooperation," in *Proc. of IEEE VTC Spring 2013*, (Dresden, Germany), June 2013.
[39] V. Jungnickel, K. Manolakis, W. Zirwas, B. Panzner, V. Braun, M. Lossow, M. Sternad, R. Apelfröjd, and T. Svensson, "The role of small cells, coordinated multi-point and massive MIMO in 5G," *IEEE Com. Mag.*, vol. 52, pp. 44–51, May 2014.
[40] R. H. Clarke and W. L. Khoo, "3-D mobile radio channel statistics," *IEEE Trans. Veh. Technol.*, vol. 46, pp. 798 – 799, 1997.
[41] E. T. Jaynes, *Probability theory: The logic of science*. Cambridge University press, 2003.
[42] P. Stoica and T. Söderström, *System Identification*. Prentice Hall, 2001.
[43] I. Wiklund, "Channel prediction for moving relays," Master's thesis, Uppsala University, 2013.
[44] G. Auer, "Pilot-symbol aided channel estimation by wiener filtering for OFDM systems with multiple transmit antennas," in *5:th IEEE Intern. Conf. on 3G Mob. Com. Tech. 2004*, 2004.
[45] D. Aronsson and M. Sternad, "OFDMA uplink channel prediction to enable frequency-adaptive multiuser scheduling," in *EUSIPCO 2007*, (Poznan, Poland), Sept. 2007.
[46] R. Fritzsche, E. Ohlmer, and G. Fettweis, "Where to predict the channel in cooperative cellular networks with backhaul delays," in *Proc. of the 9th International ITG Conference on Systems, Communications and Coding (SCC'13)*, (Munich, Germany), Jan. 2013.
[47] T. Eriksson and T. Ottosson, "Compression of feedback for adaptive transmission and scheduling," *Proceedings of the IEEE*, vol. 95, pp. 2314 – 2321, 2007.
[48] T. Wild, C. Hoek, G. Herzog, and J. Koppenborg, "Multi-antenna OFDM channel feedback compression exploiting sparsity," in *Proc. of European Wireless Conference 2013*, (Guildford, UK), Apr. 2013.
[49] G. Wunder, H. Boche, T. Strohme, and P. Jung, "Sparse signal processing concepts for efficient 5G system design," *IEEE Access*, vol. 3, pp. 195 – 208, 2013.
[50] N. Seifi, M. Viberg, R. W. Heath, J. Zhang, and M. Coldrey, "Multimode transmission in network mimo downlink with incomplete csi," *EURASIP Jour. on Advances in Signal Processing*, 2011.
[51] Q. H. Spencer, C. B. Peel, A. L. Swindlehurst, and M. Haardt, "An introduction to the multi-user MIMO downlink," *IEEE Com. Mag.*, vol. 42, pp. 60–67, 2004.
[52] R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proc. of IEEE ICC 1995*, (Seattle,

USA), 1995.
[53] T. Marzetta, "How much training is required for multiuser MIMO?," in *14th Asilomar Conf. on Sig., Sys. and Comp.*, (Pacific Grove, USA), Oct. 2006.
[54] T. Marzetta, "Massive MIMO: An introduction," *Bell Labs Tech. Jour.*, vol. 20, pp. 11–22, 2015.
[55] E. Björnsson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive mimo systems with non-ideal hardware: Energy efficiency, estimation and capacity limits," *IEEE Trans. on Inf. Theory*, vol. 60, pp. 7112–7139, Nov. 2014.
[56] E. Björnsson, E. Larsson, and T. Marzetta, "Massive MIMO: Ten myths and one critical question," *IEEE Commun. Mag.*, vol. 54, pp. 114 – 123, Feb. 2016.
[57] S. Shamai and B. M. Zaidel, "Enhancing the cellular downlink capacity via co-processing at the transmitting end," in *Proc. of IEEE VTC 2001*, (Rhodes, Greece), May 2001.
[58] S. A. Jafar and A. J. Goldsmith, "Transmitter optimization for multiple antenna cellular systems," in *Proc. of IEEE International Symposium on Inf. Theory*, 2002.
[59] D. Gesbert, S. Hanly, H. Huang, S. S. Shitz, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: a new look at interference," *IEEE Jour. on Select. Areas Com.*, vol. 28, pp. 1380–1408, 2010.
[60] Z. Ma, Z. Q. Zhang, Z. G. Ding, P. Z. Fan, and H. C. Li, "Key techniques for 5G wireless communications: network architecture, physical layer, and MAC layer perspectives," *Sci. China Inf. Sci.*, vol. 58, 2015.
[61] Ericsson AB, "5G energy performance," *Ericsson White Papers*, Apr. 2015.
[62] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE Jour. of Sel. Topics in Sig. Proc.*, vol. 8, pp. 742–758, 2014.
[63] E. Björnsson, E. Larsson, and M. Debbah, "Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated?," *IEEE Trans. Wireless Commun.*, vol. 15, pp. 1293–1308, 2016.
[64] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links," *IEEE Trans. on Inf. Theory*, vol. 49, pp. 951–963, 2003.
[65] J.-C. Guey and L. D. Larsson, "Modeling and evaluation of MIMO systems exploiting channel reciprocity in TDD mode," in *Proc. of IEEE VTC Fall 2004*, (Los Angeles, USA), Sept. 2004.
[66] J. Flordelis, F. Rusek, F. Tufvesson, E. G. Larsson, and O. Edfors, "Massive MIMO performance - TDD versus FDD: What do measurements say?," *arXiv:1704.00623*, 2017.
[67] Ericsson AB, "5G radio access," *Ericsson White Papers*, Apr. 2016.
[68] Fantastic-5G project report D4.2 , "Final results for the flexible 5G air interface multi-node/multi-antenna solution," 2017.
[69] W. Zirwas, M. Sternad, and R. Apelfröjd, "Key solutions for a massive mimo fdd system," in *Proc. of IEEE PIMRC 17*, (Montreal, Canada), Sept. 2017.
[70] R. S. Ganesan, W. Zirwas, B. Panzner, K. I. Pedersen, and K. Valkealahti, "Integrating 3D channel model and grid of beams for 5G mMIMO system level simulations," in *Proc. of IEEE VTC Fall 2016*, (Montreal, Canada), Sept. 2016.
[71] J. Choi, D. J. Love, and P. Bidigare, "Downlink training techniques for FDD

massive MIMO system: Open-loop and closed-loop training with memory," *IEEE Jour. of Sel. Topics in Sig. Proc.*, vol. 8, pp. 802 – 814, 2013.
[72] S. Noh, M. D. Zolowski, Y. Sung, and D. J. Love, "Pilot beam pattern design for channel estimation in massive MIMO systems," *IEEE Jour. on Sel. Topics in Sign. Proc.*, vol. 8, pp. 787–801, 2014.
[73] Y. Han, J. Lee, and D. J. Love, "Compressed sensing-aided downlink channel training for massive MIMO systems," *IEEE Trans. on Com.*, 2017. In press, DOI 10.1109/TCOMM.2017.2691700.
[74] Z. Jiang, A. F. Molisch, G. Caire, and Z. Niu, "On the achievable rates of FDD massive MIMO systems with spatial channel correlation," in *Proc. of IEEE ICCC 2014*, (Shanghai, China), Oct. 2014.
[75] C. C. Tseng, J. Y. Wu, and T. S. Lee, "Enhanced compressive downlink CSI recovery for FDD massive MIMO systems using weighted block l_1-minimization," *IEEE Trans. on Com.*, vol. 64, pp. 1055–1067, 2016.
[76] Z. Gao, L. Dai, W. Dai, B. Shim, and Z. Wang, "Structured compressive sensing based spatio-temporal joint channel estimation for FDD massive MIMO," *IEEE Trans. on Com.*, vol. 64, pp. 601–617, 2015.
[77] B. Tomasi, A. Decurninge, and M. Guillaud, "SNOPS: Short non-orthogonal pilot sequences for downlink channel state estimation in FDD massive MIMO," in *IEEE Globecom Works.*, (Washington DC, USA), Dec. 2016.
[78] M. Kurras, L. Thiele, T. Haustein, and C. Y. W. Lei, "Full dimension MIMO for frequency division duplex under signaling and feedback constraints," in *European Sig. Proc. Conf. (EUSIPCO)*, (Budapest, Hungary), Aug./Sept. 2016.
[79] H. Tröger, T. Weber, M. Meurer, and P. W. Baier, "A novel performance evaluation technique for joint transmission multiuser downlinks with multi-element transmit antennas," in *Proc. of the COST 262 Works.*, (Schloss Reisensburg, Germany), Jan. 2001.
[80] J. Bartelt, P. Rost, D. Wubben, J. Lessmann, B. Melis, and G. Fettweis, "Fronthaul and backhaul requirements of flexibly centralized radio access networks," *IEEE Wireless Commun.*, vol. 22, pp. 105 – 111, 2015.
[81] X. Tao, X. Xu, and Q. Cui, "An overview of cooperative communications," *IEEE Wireless Com. Mag.*, vol. 8, pp. 65–171, 2012.
[82] D. Lee, H. Seo, B. Clerckx, E. Hardouin, D. Mazzarese, S. Nagata, and K. Sayana, "Coordinated multipoint transmission and reception in LTE-advanced deployment: Scenarios and operational challenges," *IEEE Wireless Com. Mag.*, vol. 50, pp. 148–155, 2012.
[83] E. Björnsson, N. Jaldén, M. Bengtsson, and B. Ottersten, "Optimality properties, distributed strategies, and measurement-based evaluation of coordinated multicell OFDMA transmission," *IEEE Trans. Wireless Commun.*, vol. 15, pp. 1293–1308, Feb. 2016.
[84] 3GPP TR 36.814 v9.0.0, "Further advancements for E-UTRA physical layer aspects (release 9)," 2010.
[85] T. Wild, "Comparing downlink coordinated multi-point schemes with imperfect channel knowledge," in *Proc. of IEEE VTC fall 2011*, (San Francisco, USA), 2011.
[86] 3GPP TR 36.819 v11.0.0, "Technical specification group radio access network; Coordinated multi-point operation for LTE physical layer aspects,

(release 11)," 2011.
[87] W. Lee, I. Lee, J. S. Kwak, B. Ihm, and S. Han, "Multi-BS MIMO cooperation: challenges and practical solutions in 4G systems," *IEEE Wireless Com.*, vol. 19, pp. 89–96, 2012.
[88] L. Su, C. Yang, and S. Han, "The value of channel prediction in CoMP systems with large backhaul latency," in *Proc. of IEEE WCNC 2012*, (Paris, France), Apr. 2012.
[89] T. Ekman, G. Kubin, M. Sternad, and A. Ahlen, "Quadratic and linear filters for radio channel prediction," in *Proc. of IEEE VTC Fall 1999*, (Amsterdam, Netherlands), Sept. 1999.
[90] J. Lee, Y. Kim, H. Lee, B. L. Ng, D. Mazzarese, J. Liu, W. Xiao, and Y. Zhou, "Coordinated multipoint transmission and reception in LTE-advanced systems," *IEEE Com. Mag.*, vol. 50, pp. 44–55, 2012.
[91] M. K. Karakayali, G. J. Foschini, and R. A. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *IEEE Wireless Com.*, vol. 13, pp. 56–61, 2006.
[92] J. Zhang, R. Chen, J. G. Andrews, A. Ghosh, and R. W. Heath, "Network MIMO with clustered linear precoding," *IEEE Trans. Wireless Com.*, vol. 8, pp. 1910–1921, 2009.
[93] H. Zhang and H. Dai, "Cochannel interference mitigation and cooperative processing in downlink multicell multiuser MIMO networks," *EURASIP Jour. on Wirel. Com. and Netw.*, 2004.
[94] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H.-P. Mayer, L. Thiele, and V. Jungnickel, "Coordinated multipoint: Concepts, performance and field trial results," *IEEE Com. Magazine*, vol. 49, pp. 102–111, 2011.
[95] M. Grieger and G. Fettweis, "Large scale field trial results on time domain compression for uplink joint detection," in *Proc. of IEEE PIMRC 2011*, (Toronto, Canada), Sept. 2011.
[96] W. Mennerich and W. Zirwas, "Reporting effort for cooperative systems applying interference floor shaping," in *Proc. of IEEE PIMRC 2011*, (Toronto, Canada), Sept. 2011.
[97] H. Huh, G. Caire, H. C. Papadopoulos, and S. A. Ramprashad, "Achieving "massive MIMO" spectral efficiency with a not-so-large number of antennas," *IEEE Trans. Wireless Commun.*, vol. 11, pp. 3226–3239, 2012.
[98] A. Lozano, R. W. Heath, and J. G. Andrews, "Fundamental limits of cooperation," *IEEE Trans. on Inf. Theory*, vol. 59, pp. 5213–5226, 2013.
[99] W. Mennerich, M. Grieger, W. Zirwas, and G. Fettweis, "Interference mitigation framework for cellular mobile radio networks," *International Journal of Antennas and Propagation*, 2013.
[100] L. Thiele, M. Olbrich, M. Kurras, and B. Matthiesen, "Channel aging effects in CoMP transmission: Gains from linear channel prediction," in *45th Asilomar Conf. on Sig., Sys. and Comp.*, (Pacific Grove, USA), Nov. 2011.
[101] J. Lee and N. Jindal, "Dirty paper coding vs. linear precoding for MIMO broadcast channels," in *Proc. of IEEE ACSSSC'06*, (Pacific Grove, USA), Oct. 2006.
[102] K. Öhrn, A. Ahlen, and M. Sternad, "A probabilistic approach to multivariable

robust filtering and open-loop control," *IEEE Trans. on Automatic Control*, vol. 40, pp. 405–418, 1995.
[103] L.-J. Bränmark, *Robust sound field control for audio reproduction - A polynomial approach to discrete-time acoustic modeling and filter design*. PhD thesis, Uppsala University, Jan. 2011.
[104] R. Hunger, F. A. Dietrich, M. Joham, and W. Utschick, "Robust transmit zero-forcing filter," in *Proc. of ITG Works. on Smart Antennas*, (Munich, Germany), Mar. 2004.
[105] P. M. Castro, M. Joham, L. Castedo, and W. Utschick, "Robust MMSE linear precoding for multiuser miso systems with limited feedback and channel prediction," in *Proc. of IEEE VTC Fall 2008*, (Cannes, France), Sept. 2008.
[106] M. B. Shenouda and T. N. Davidson, "On the design of linear transceivers for multiuser systems with channel uncertainty," *IEEE Jour. on Sel. Areas in Com.*, vol. 26, pp. 1015–1024, 2008.
[107] X. Zhang, D. P. Palomar, and B. Ottersten, "Statistically robust design of linear MIMO transceivers," *IEEE Trans. Signal Process.*, vol. 56, pp. 3678–3689, 2008.
[108] R. Fritzsche and G. Fettweis, "Robust precoding with general power constraints considering unbounded channel uncertainty," in *Proc. of IEEE ISWCS 2012*, (Paris, France), Aug. 2012.
[109] T. R. Lakshmana, R. Apelfröjd, T. Svensson, and M. Sternad, "Particle swarm optimization based precoder in comp with measurement data," in *5th systems and network optimization for wirelsess (SNOW) workshop*, (Åre, Sweden), Apr. 2014.
[110] T. R. Lakshmana, C. Botella, and T. Svensson, "Partial joint processing with efficient backhauling using particle swarm optimization," *EURASIP Jour. on Wireless Com. and Netw.*, 2012.
[111] P. Marsch and G. Fettweis, "On multicell cooperative transmission in backhaul constrained cellular systems," *Ann. Telecommun.*, vol. 63, pp. 253–269, 2008.
[112] A. Papadogiannis, H. Bang, D. Gesbert, and E. Hardouin, "Efficient selective feedback design for multicell cooperative networks," *IEEE Trans. on Vehicular Technology*, vol. 60, pp. 196–205, 2011.
[113] C. T. K. Ng and H. Huang, "Linear precoding in cooperative MIMO cellular networks with limited coordination clusters," *IEEE Jour. on Sel. Areas in Com.*, vol. 28, pp. 1446 – 1454, 2010.
[114] T. R. Lakshmana, A. Tölli, R. Devassy, and T. Svensson, "Precoder design with incomplete feedback for joint transmission," *IEEE Trans. Wireless Commun.*, vol. 15, pp. 1923 – 1936, 2016.
[115] P. Kerret and D. Gesbert, "Sparse precoding in multicell MIMO systems," in *Proc. of IEEE WCNC 2012*, (Paris, France), Apr. 2012.
[116] R. Y. Chang, Z. Tao, J. Zhang, and C. C. J. Kuo, "Multicell OFDMA downlink resource allocation using a graphic framework," *IEEE Trans. on Vehicular Technology*, vol. 58, pp. 3494–3507, 2009.
[117] Z. Shen, R. Chen, J. G. Andrews, R. W. Heath, and B. L. Evans, "Low complexity user selection algorithms for multiuser MIMO systems with block diagonalization," *IEEE Trans. Signal Process.*, vol. 54, pp. 3658–3663, 2006.
[118] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-antenna downlink channels with

limited feedback and user selection," *IEEE Jour. on Sel. Areas in Com.*, vol. 25, pp. 1478–1491, 2007.
[119] F. Diehm and G. Fettweis, "Centralized scheduling for joint decoding cooperative networks subject to signaling delays," in *Proc. of IEEE VTC Fall 2011*, (San Francisco, USA), 2011.
[120] A. Klockar, M. Sternad, A. Brunström, and R. Apelfröjd, "User-centric pre-selection and scheduling for coordinated multipoint systems," in *Proc. of IEEE ISWCS14*, (Barcelona, Spain), Aug. 2014.
[121] P. Viswanath, D. N. C. Tse, and R. L. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inf. Theory*, vol. 42, pp. 1277–1294, 2002.
[122] T. Bonald, "A score-based opportunistic scheduler for fading radio channels," in *Proc. of European Wireless Conf (EWC)*, (Barcelona, Spain), Feb. 2004.
[123] J. Medbo, I. Siomina, A. Kangas, and J. Furuskog, "Propagation channel impact on LTE positioning accuracy - A study based on real measurements of observed time difference of arrival," in *Proc. of IEEE PIMRC 2009*, (Tokyo, Japan), Sept. 2009.
[124] Y. Sui, A. Papadogiannis, and T. Svensson, "The potential of moving relays - a performance analysis," in *Proc. of IEEE VTC Spring 2012*, (Yokohama, Japan), May 2012.
[125] W. Zirwas and M. Haardt, "Channel prediction for B4G radio systems," in *Proc. of IEEE VTC Spring 2013*, (Dresden, Germany), June 2013.
[126] A. B. Martinez, "Using predictor antennas for long-range prediction of fast fading for moving relays," Master's thesis, Technische Universität Dresden, 2012.
[127] N. Jamaly, R. Apelfröjd, A. B. Martinez, M. Grieger, T. Svensson, M. Sternad, and G. Fettweis, "Analysis and measurement of multiple antenna systems for fading channel prediction in moving relays," in *8th European conference on antennas and propagation (EuCAP)*, (Hague, Netherlands), Apr. 2014.
[128] J. Björsell, M. Sternad, and M. Grieger, "Using predictor antennas for the prediction of small-scale fading provides an order-of-magnitude improvement of prediction horizons," in *Proc. of IEEE ICC 17*, (Paris. France), July 2017.
[129] J. Björsell, M. Sternad, and M. Grieger, "Predictor antennas in action," in *Proc. of IEEE PIMRC 17*, (Montreal, Canada), Oct. 2017.
[130] D.-T. Phan-Huy, S. Wesemann, J. Björsell, and M. Sternad, "Adaptive massive MIMO for fast moving connected vehicles: It will work with predictor antennas!," in *22nd International Workshop on Smart Antennas (WSA2018)*, (Bochum, Germany), March 2018.
[131] R. G. Vaughan, N. L. Scott, and J. P. T. T. K. Southon, "Characterization of standing wave movement by in-line antennas," in *Proc. of Antennas and Propagation Society Symposium 1991 Digest*, (London, Canada), June 1991.
[132] R. G. Vaughan and N. L. Scott, "Closely spaced terminated monopoles for vehicular diversity antennas," in *Proc. of Antennas and Propagation Society Symposium 1992 Digest*, (Chicago, USA), June 1992.
[133] D.-T. Phan-Huy, M. Sternad, and T. Svensson, "Making 5G adaptive antennas work for very fast moving vehicles," *IEEE Intelligent Transportation System Magazine*, pp. 71 – 84, 2015.
[134] X. Lu, H. Wang, and M. Li, "Kalman fixed-interval and fixed-lag smoothing

for wireless sensor systems with multiplicative noises," in *Proc. of IEEE CCDC 2012*, (Taiyuan, China), July 2012.
[135] S. Park and S. Choi, "Iterative equalizer based on kalman filtering and smoothing for MIMO-ISI channels," *IEEE Trans. Signal Process.*, vol. 63, pp. 5111–5120, 2015.

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 1642

Editor: The Dean of the Faculty of Science and Technology

A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-344270

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2018