Toni Levanen

Superimposed Training for Single Carrier Transmission in Future Mobile Communications

Tampere 2014

Toni Levanen

Superimposed Training for Single Carrier Transmission in Future Mobile Communications

Thesis for the degree of Doctor of Science in Technology to be presented with due permission for public examination and criticism in Tietotalo Building, Auditorium TB109, at Tampere University of Technology, on the 29th of August 2014, at 12 noon.

Supervisor

Markku Renfors, Dr. Tech., Professor
Department of Electronics and Communications Engineering
Tampere University of Technology
Tampere, Finland

Pre-examiners

Desmond McLernon, Ph.D., Reader
School of Electronic and Electrical Engineering
University of Leeds
Leeds, United Kingdom

Kimmo Kansanen, Dr. Tech., Professor
Norwegian University of Science and Technology
Trondheim, Norway

Opponent

Mounir Ghogho, Ph.D., Professor
School of Electronic and Electrical Engineering
University of Leeds
Leeds, United Kingdom

ABSTRACT

The amount of wireless devices and wireless traffic has been increasing exponentially for the last ten years. It is forecasted that the exponential growth will continue without saturation till 2020 and probably further. So far, network vendors and operators have tackled the problem by introducing new evolutions of cellular macro networks, where each evolution has increased the physical layer spectral efficiency. Unfortunately, the spectral efficiency of the physical layer is achieving the Shannon-Hartley limit and does not provide much room for improvement anymore.

However, considering the overhead due to synchronization and channel estimation reference symbols in the context of physical layer spectral efficiency, we believe that there is room for improvement. In this thesis, we will study the potentiality of superimposed training methods, especially data-dependent superimposed training, to boost the spectral efficiency of wideband single carrier communications even further.

The main idea is that with superimposed training we can transmit more data symbols in the same time duration as compared to traditional time domain multiplexed training. In theory, more data symbols means more data bits which indicates higher throughput for the end user. In practice, nothing is free. With superimposed training we encounter self-interference between the training signal and the data signal. Therefore, we have to look for iterative receiver structures to separate these two or to estimate both, the desired data signal and the interfering component.

In this thesis, we initiate the studies to find out if we truly can improve the existing systems by introducing the superimposed training scheme. We show that in certain scenarios we can achieve higher spectral efficiency, which maps directly to higher user throughput, but with the cost of higher signal processing burden in the receiver. In addition, we provide analytical tools for estimating the symbol or bit error ratio in the receiver with a given parametrization.

The discussion leads us to the conclusion that there still remains several open topics for further study when looking for new ways of optimizing the overhead of reference symbols in wireless communications. Superimposed training with data-dependent

components may prove to provide extra throughput gain. Furthermore, the superimposed component may be used for, e.g., improved synchronization, low bit-rate signaling or continuous tracking of neighbor cells. We believe that the current systems could be improved by using the superimposed training collectively with time domain multiplexed training.

PREFACE

Research work that led to this dissertation has been carried out in the Department of Electronics and Communications Engineering (ELT), Tampere University of Technology (TUT), Finland, during the years 2007-2014.

The work presented in this dissertation would not have been possible without the support of numerous people. First of all, I would like to thank my supervisor, Professor Markku Renfors, for his thoughtful guidance and positive feedback over the years I have spent on completing this dissertation. Together with Professor Mikko Valkama, I would like to thank them for giving me the possibility to work in our Department, allowing me to participate in various research projects and supporting me throughout my studies.

I am grateful to the reviewers of this dissertation, Dr. Des McLernon and Dr. Kimmo Kansanen, for agreeing to review my thesis and providing constructive comments to improve the overall quality of the thesis. I would also like to thank Prof. Mounir Ghogho for agreeing to act as opponent in the public examination of my dissertation.

One of the main reasons why I have enjoyed working in the Department of Electronics and Communications Engineering, is all the people working there, building a great working environment and atmosphere for scientific research. Especially, I would like to thank my colleagues Tero Ihalainen, Jukka Talvitie, Markus Allén, Jaakko Marttila, Danai Skournetou, Tero Isotalo, Tobias Hidalgo Stitz, Ali Shahed, Stanislav Nonchev, Yaning Zou, Ville Syrjälä, Lauri Anttila, Vesa Lehtinen, Jussi Turkka, Ari Asp, and Aki Hakkarainen for their support and numerous discussions on more or less scientific topics. When it comes to practical matters, I have to thank especially Tarja Erälaukko for looking over us, and also Sari Kinnari, Kirsi Viitanen, and Marianna Jokila for their hard work on keeping the big wheels of science turning in our department. In addition, I would like to thank Pertti Koivisto and Antti Niemistö for their hard work in TISE.

Finally, I would like to thank my family for their unconditional love and for supporting me in my studies over all these years. Especially, my mother Rauni and father Raimo for making this possible, my elder brothers Tapani and Tapio for teaching me also other aspects of life, and my little sister Janita for believing in me. Also, the support of all of my friends not mentioned here should not be forgotten. Most of all, I thank my beloved Riitta and our beautiful daughter Mila for providing me love, strength, understanding, and courage to continue, no matter how desperate things might have looked like. This dissertation would not have become reality without these two angels in my life. Thank you for teaching me what really is important in life.

> *You need to study for life, but don't live for studies,*
> *you need to work to live, but don't live for work.*
> *You learn life through loving and being loved,*
> *you learn life by falling and standing up,*
> *you learn life by fighting and forgiving.*
> *When you learn that you are bigger together than alone,*
> *when you realize that the most important thing in life is now and here,*
> *you have taken the first step to understand how to live your life.*

Tampere, August 2014,

Toni Levanen

TABLE OF CONTENTS

LIST OF PUBLICATIONS

This thesis consists of five publications, which in the text are referred to as Publications [P1]-[P5]. Publications [P1],[P2] focus on improving the channel estimation analysis for superimposed training and the last three publications [P3]-[P5] are related to the research on data-dependent superimposed pilots in single-input multiple-output and multiple-input multiple-output communications. In publications [P3],[P4], the peak-to-average-power ratio increase caused by data-dependent superimposed training is studied and methods to reduce the effect are proposed. In publication [P5], the energy efficiency of data-dependent superimposed training in multiple-input multiple-output communications is studied and performance comparisons with time domain multiplexed training are provided. Publication [P4] is an article published in a journal, while the rest are conference publications.

[P1] T. Levanen, J. Talvitie, and M. Renfors, "Improved Performance Analysis for Superimposed Pilot Based Short Channel Estimator" in *The 11th International Wrokshop on Signal Processing Advances in Wireless Communications, SPAWC 2010*, Marrakech, Morocco, 20-23 June 2010.

[P2] T. Levanen and M. Renfors, "Improved Performance Bounds for Iterative IC LMMSE Channel Estimator with SI Pilots", in *The 21st Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2010*, Istanbul, Turkey, 26-29 Sept. 2010.

[P3] T. Levanen, J. Talvitie, and M. Renfors, "Performance Evaluation of a DDST Based SIMO SC System with PAPR Reduction" in *The 6th International Symposium on Turbo Codes & Iterative Information Processing, ISTC 2010*, Brest, France, 6-10 Sept. 2010.

[P4] T. Levanen, J. Talvitie, and M. Renfors, "Throughput Evaluation of Time-Multiplexed and Data-Dependent Superimposed Training Based Transmission with Practical Power Amplifier Model" in *Eurasip Journal on Wireless Communications and Networking* special issue on *Algorithm and Implementation*

Aspects of Channel Codes and Iterative Receivers, Vol. 49, 2012, doi:10.1186/1687-1499-2012-49.

[P5] T. Levanen and M. Renfors, "Improved Energy Efficiency for Wireless SC MIMO Through Data-Dependent Superimposed Training", in *the 5th International Symposium on Communications, Control, and Signal Processing, IS-CCSP 2012*, Rome, Italy, 2-4 May 2012.

LIST OF ABBREVIATIONS

List of the most important and recurrently appearing abbreviations in this thesis.

3GPP	3rd Generation Partnership Project
ADC	Analog-to-Digital Converter
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BPSK	Binary Phase Shift Keying
CCDF	Complementary Cumulative Distribution Function
CM	Cyclic Mean
CP	Cyclic Prefix
DDST	Data-Dependent Superimposed Training
DFT	Discrete Fourier Transform
E_b/N_0	Energy per bit to single sided noise spectral density
FB	Filter Bank
FDE	Frequency-Domain Equalization
IDFT	Inverse Discrete Fourier Transform
IPI	Inter-Packet Interference
LMMSE	Linear Minimum Mean Squared Error
LS	Least Squares
LTE	Long Term Evolution
MCS	Modulation and Coding Scheme
MIMO	Multiple-Input Multiple-Output
ML	Maximum Likelihood
MSE	Mean Squared Error
OFDM	Orthogonal Frequency Division Multiplexing
PA	Power Amplifier
PAM	Pulse Amplitude Modulation
PAPR	Peak-to-Average-Power Ratio
PDF	Probability Density/Distribution Function

QAM   Quadrature Amplitude Modulation
QPSK   Quadrature Phase Shift Keying
RRC   Root-Raised Cosine
SC   Single Carrier
SCE   Sub-Channel wise Equalizer
SIMO   Single-Input Multiple-Output
SISO   Single-Input Single-Output
SIT   Superimposed Training
SNR   Signal to noise ratio
TDMT   Time-Domain Multiplexed Training
SSPA   Solid State Power Amplifier
WLAN   Wireless Local Area Network

LIST OF PRINCIPAL SYMBOLS

List of the principal symbols and notations used throughout this thesis.

a	Distance between a constellation point and a decision threshold
$\mathbf{c}$	Coded bit vector
$\hat{\mathbf{c}}^i$	Soft coded bit estimate vector at detection iteration i
$\hat{\mathbf{c}}_\nu^i$	Soft coded bit estimate vector related to the νth symbol
$\hat{c}_\nu^i(q)$	Soft data bit estimate (Log-likelihood presentation of the ratio of a posteriori probabilities) of the qth bit of the νth symbol at i^{th} detection iteration
$\mathbf{d}$	Data symbol vector for general derivations valid for all training methods
$\mathbf{d}_{DDST}$	Data symbol vector for DDST
$\mathbf{d}_{SIT}$	Data symbol vector for SIT
$\mathbf{d}_{TDMT}$	Data symbol vector for TDMT, which is shorter than $\mathbf{d}_{DDST}$ or $\mathbf{d}_{SIT}$ if equal total length assumed
$\hat{\mathbf{d}}^i$	Data symbol estimate at detection iteration i
$\mathbf{D}_M$	Cyclic mean of the full data symbol matrix $\mathbf{D}$
$\mathbf{h}$	Channel vector
$\hat{\mathbf{h}}$	Estimate of the channel vector
$\mathbf{H}$	Full channel matrix
$\mathbf{H}_d$	Fraction of the channel matrix with TDMT related to data symbols
$\mathbf{H}_k$	Frequency domain presentation of the channel weights at frequency bin k
$\mathbf{H}_{k,RRC}$	Frequency domain weight of the RRC pulse shaping
$\mathbf{H}_p$	Fraction of the channel matrix with TDMT related to only pilot symbols

$IDX_{pilot,t}$	Indexing set to obtain pilot vector transmitted from Tx antenna t
$IDX_{pilot,t,cp}$	Indexing set to obtain cyclic prefix for the pilot vector transmitted from Tx antenna t
$\mathbf{J}(a,b,c)$	Matrix used to evaluate a copies of cyclic mean over b cycles of length c
k	Number of symbols in a constellation
K	Overlapping factor with filter banks
L_c	Channel response length
M	Filter bank size, number of subchannels
$\mathbf{M}_{AddCP}(N, L_c - 1)$	Matrix for extending a vector of length N with a cyclic prefix of length $L_c - 1$
$\mathbf{M}_{RemoveCP}(N, L_c - 1)$	Matrix for removing a CP of length $L_c - 1$ from a vector of length $N + L_c - 1$
n	Sample index
$\mathbf{n}$	Noise vector
$\mathbf{n}_M$	Cyclic mean vector of the noise vector
N_c	Number of cyclic copies
N_{cp}	Number of samples in the cyclic prefix
N_d	Number of samples in the data field
$\mathbf{N}_k$	Frequency domain presentation of the noise samples at frequency bin k
N_p	In DDST, length of a component pilot sequence in a cyclic pilot sequence,and in TDMT, length of the pilot sequence without CP
N_{tot}	Total length of the packet, containing pilot and data fields
N_{Rx}	Number of receiveing antennaes
N_{Tx}	Number of transmitting antennaes
$\delta\mathbf{p}_{SIT}$	Residual pilot signal component in the received signal with SIT after the removal of power scaled pilot signal
$\mathbf{p}_{c,DDST}$	Cyclic pilot sequence for DDST
$\mathbf{p}_{c,SIT}$	Cyclic pilot sequence for SIT
$\mathbf{p}_d$	Data dependent pilot; self interference on top of symbol sequence $\mathbf{d}_{DDST}$ caused by the removal of the cyclic mean
$\mathbf{p}_{DDST}$	Basis pilot sequence for DDST
$\mathbf{p}_{SIT}$	Basis pilot sequence for SIT
$\mathbf{p}_{TDMT}$	Pilot sequence for TDMT

$\mathbf{P}_{DDST}$	Circulant pilot matrix used for channel estimation with DDST,
$\mathbf{P}_{SIT}$	Circulant pilot matrix used for channel estimation with SIT,
$\mathbf{P}_{TDMT}$	Circulant pilot matrix used for channel estimation with TDMT,
P_{AVG}	Average power of the transmitted signal from certain antenna element
q	Number of bits per symbol
r	Index value for receiving antenna
s	Sequence index for Zadoff-Chu sequences
t	Index for transmitting antenna
$\mathbf{w}$	Time domain channel equalizer weight matrix
$\mathbf{x}_{DDST}$	Transmitted signal for DDST without cyclic prefix
$\mathbf{x}_{DDST,cp}$	Transmitted signal for DDST with cyclic prefix
$\mathbf{x}_{SIT}$	Transmitted signal for SIT without cyclic prefix
$\mathbf{x}_{SIT,cp}$	Transmitted signal for SIT with cyclic prefix
$\mathbf{x}_t$	Signal vector transmitted from antenna index t
$\mathbf{x}_{TDMT}$	Transmitted signal for TDMT without cyclic prefix
$\mathbf{x}_{TDMT,cp}$	Transmitted signal for TDMT with cyclic prefix
$\mathbf{X}$	Full transmitted signal matrix
$\mathbf{X}_k$	Frequency domain presentation of the transmitted samples at frequency bin k
$\mathbf{y}_d$	Data portion of the received sequence with TDMT
$\mathbf{y}_{DDST}$	Received signal with DDST without cyclic prefix
$\mathbf{y}_{DDST,cp}$	Received signal with DDST with cyclic prefix
$\mathbf{y}_p$	Pilot portion of the received sequence without cyclic prefix with TDMT
$\mathbf{y}_{p,cp}$	Pilot portion of the received sequence with cyclic prefix with TDMT
$\mathbf{y}_r$	Received signal vector in antenna with index r
$\mathbf{y}_{r,M}$	Cyclic mean of the received signal vector in antenna index r
$\mathbf{y}_{SIT}$	Received signal with SIT without cyclic prefix
$\mathbf{y}_{SIT,cp}$	Received signal with SIT with cyclic prefix
$\mathbf{Y}_k$	Frequency domain presentation of the received samples at frequency bin k

α	Scaling factor for the input signal in Bussgang's theorem
β	Power scaling factor for DDST and SIT to allocate data signal power
	for cyclic prefix
γ	Fraction of total data signal power allocated for pilots in all training schemes
Δ	Channel estimation error matrix
$\Delta\mathbf{p}_{SIT}$	Residual pilot error term with SIT after pilot removal
ε	Estimation error term
ζ	Distance between two constellation points
η	Integer number required for Zadoff-Chu sequence generation
ι	Sample index with range $[0, 1, \ldots, N_c - 1]$
κ	Sample index with range $[0, 1, \ldots, N_p - 1]$
λ	Power scaling term for TDMT pilot sequence when fraction of γ
	of the total data signal power is allocated for pilots
ν	Data symbol index
ξ	Power value indicated in the x-axis in figures plotting complementary cumulative distributions for peak-power and peak-to-averate-power ratio
ρ	RRC filter rolloff factor
σ_x	Standard deviation of random variable x
Ω_D	Set of symbols, models the used constellation
Ω_Z	Set of (cyclic) mean values over the N_c symbols from the symbol set Ω_D
$\mathbf{1}_{a \times b}$	Matrix of all ones with a rows and b columns
$\delta(x)$	Delta function, which has value of 1 if $x = 0$, and 0 otherwise
$\text{diag}(\cdot)$	For matrices gives the main diagonal values and for vectors generates a diagonal matrix with vector elements on the main diagonal
$\text{E}(x)$	Expectation operation on x over the distribution of $p(x)$ or $f(x)$ for discontinuous or continuous case respectively
$f(x)$	Probability density function for continuous random variable x
$G(x)$	Power amplifier model

$\text{mod}(a,b)$	Modulus operator that gives a positive integer remainder of the Euclidean division of a by b
$\mathbf{I}_a$	Square identity matrix with a columns and rows
$p(x)$	Probability distribution function for discrete random variable x
$p(a|b)$	Probability of a with the condition of b
$P_{app}(x)$	A posteriori probability of x
$\mathbf{S}(\cdot)$	Predistortion model
$u(t)$	Step function, $u(t) = 1$ if $t \geq 0$, and $u(t) = 0$ otherwise
$(\cdot)^H$	Hermitian transpose
$(\cdot)^T$	Transpose
$\mathbf{A} \propto (a,b)$	Used to indicate the dimensions of a matrix or a vector (matrix $\mathbf{A}$ has a rows and b columns)
$\mathbf{a} \odot \mathbf{b}$	Element-wise product between vectors $\mathbf{a}$ and $\mathbf{b}$
$\otimes(\mathbf{A}, \mathbf{B})$	Kronecker product over matrices $\mathbf{A}$ and $\mathbf{B}$
$|\Omega|$	Cardinality of the set Ω
$\|\mathbf{a}\|$	Vector norm for the vector $\mathbf{a}$

1. INTRODUCTION

1.1 Background and Motivation

Wireless communication systems have evolved and conquered the world during the last 20 years. No one could predict the speed and impact at which they changed our world and how we perceive our environment, connect with our friend around the world, and share information any time, anywhere.

Due to the way we have adapted our way of living with the wireless communication systems, we drive the exponential growth of capacity required from wireless systems. The recent advancements in waveform design and channel coding have significantly increased the physical layer throughput. However, there is a limit for these improvements that we can not exceed. Thus, an obvious step to improve the throughput experienced by the user is to look at the different overheads in the physical layer signal. By overhead we mean all the information exchanged between transmitter and receiver other than the user specific data.

Significant part of this overhead is caused by the higher layers of the wireless communication system in consideration, but we do not address this overhead in this dissertation. It is for other researchers and Ph.D. students to solve. We have concentrated on the overhead caused by the traditional time domain multiplexed training (TDMT) signal, that is used to estimate the channel between the transmitter and receiver. Also, synchronization in time and frequency domain is required for receiver and is typically achieved based on the same TDMT signal, but we have not covered synchronization aspects in this dissertation.

The problem with TDMT is that the training signal directly reduces the user data rate by reducing the number of data carrying symbols in a frame. An alternative approach to include training signal in the transmitted frame is to directly add it on top of the data symbols. This is called superimposed training (SIT). In [7, 23, 32, 33, 50, 51, 72, 74, 87, 106, 107, 110, 116] and references therein the different aspects of SIT are discussed in more detail. Additional information can be found, e.g., in [6, 19, 37,

38, 59, 65, 75, 79, 86, 95, 113, 114, 125, 126]. The main idea is that through iterative signal processing in the receiver, combining the knowledge of training signal and data symbol estimates, we could achieve similar or better throughput when compared to TDMT.

The main problem with SIT is the large self-interference from the user data symbols to the training signal. This interference typically reduces the channel estimation accuracy to the level that the detection of the data signal is infeasible even with iterative reception algorithms. A candidate solution to this is the data-dependent superimposed training (DDST), see for example [8, 45, 46, 68, 76, 81] and references therein.

In DDST, we somehow remove the self-interference term in the transmitter, allowing the receiver to obtain self-interference free channel estimates. The drawback is that now the data symbols are distorted by the removal of the self-interference. Our belief is that because the data signal is built from finite, discrete set of symbols it is easier to iteratively reconstruct the data signal in the receiver given the good channel estimate obtained with DDST.

The DDST scheme, in which we will concentrate in this thesis, is based on removal of the cyclic mean of the data signal and adding cyclic pilot sequence on top of the predistorted data signal. The DDST concept can be looked from two different aspects. First of all, if we consider that the training signal is composed of the cyclic pilot signal and removed cyclic mean together, it can be considered as data-dependent training sequence, because it depends on the data vector. On the other hand, if we consider the removed cyclic mean component to be part of the modulation, we can look at the DDST concept as time varying modulation, where the modulation is different for each data vector.

1.2   Scope and Objectives of Research

The main objective of our research is to find out whether we can improve the spectral efficiency or reduce the required energy-per-bit with DDST when compared to TDMT. We are not that interested in traditional bit error rate (BER) versus signal-to-noise ratio (SNR) results. Instead we rather compare the BER versus energy-per-bit over one-sided-noise-spectral-density (E_b/N_0) or look at the number of correctly received user data packets per second versus E_b/N_0, that defines the throughput of the system. In addition, the iterative receiver structures required for efficient detection of signals with SIT or DDST are of great interest.

In all of the publications, we have assumed that the synchronization in time and frequency is perfect. This is a rather strong assumption, and future studies including synchronization aspects are of great interest. Especially, the differences in the frame structures with TDMT and DDST and their implications need to be addressed. Based on our understanding, TDMT requires more accurate time synchronization to correctly separate pilot and data sections. For DDST, it should be sufficient to have synchronization accuracy inside the cyclic prefix and let later signal processing find the proper circular shift for correct detection. A simplified description of the main differences between different multiplexing modes is provided in Appendix A.

Also, in all our publications a filter bank (FB) based linear minimum mean squared error (LMMSE) frequency domain channel equalizer [55] was used. This structure was chosen in the beginning of the studies because of its excellent spectral containment properties and performance approaching ideal fractionally-spaced linear equalizer. Frequency domain channel equalizers are the only viable solution for wideband SC communications, in order to have computationally tractable solution. In Appendix B we provide an introduction to FB based structures and in Appendix C the FB based channel equalization is described in more detail.

Wideband SC communications system was chosen because it has received less attention after the general acceptance of the orthogonal frequency division multiplexing (OFDM). Many modern communication systems rely on OFDM, like the IEEE wireless local area network (WLAN) 802.11 family [57] and the 3rd generation partnership project (3GPP) evolved-universal terrestrial radio access (E-UTRA) physical layer, commonly referred as long term evolution (LTE) [3] downlink. Nonetheless, the SC transmission has its place and currently there is a growing interest towards it as a 5G waveform [11] for millimeter-wave communications [48]. Some additional comments and comparison between wideband SC and multicarrier communications are provided in Appendix D.

The main problem with OFDM transmission is the large peak-to-average-power ratio (PAPR) caused by summation of large number of independent active subcarriers. Having a large PAPR indicates that the used power amplifiers (PAs) have to be very linear meaning higher costs, or that the used power backoff has to be larger meaning lower radiated power. For a mobile device neither of these options is desired.

Therefore, LTE uplink uses SC frequency domain multiple access (SC-FDMA). In SC-FDMA, the SC transmission of single user is transformed to the frequency domain through critically sampled discrete Fourier transform (DFT). Then, the FFT-presentation of the signal is extended by zero samples in a process called subcarrier

mapping, either by keeping a continuous allocation or by spreading the signal to the frequency domain. Finally, an inverse DFT (IDFT) is applied to the extended signal providing us a SC like signal with lower PAPR than in OFDM signal. See [28] for discussion and comparisons between OFDMA and SC-FDMA.

In IEEE 802.11ad [58], the SC mode is used as the robust mode, providing longer link distances but supporting lower peak rates than the OFDM based alternative. Recently, in [48] a null-prefixed SC waveform was proposed for millimeter-wave [92] communications. Thus, there is again good momentum in the SC oriented research for 5G communications [11]. Furthermore, old and well optimized SC based solutions, like global system for mobile communications (GSM) [2], are again considered for device-to-device (D2D) communications [67].

1.3   Outline and Main Results of the Thesis

The main emphasis in the first publications [P1],[P2] was on the channel estimation performance with truncated channel estimators, which directly maps to symbol error probability. By truncated channel estimators we mean estimators whose length is smaller than the maximum length of the effective channel, including transmitter and receiver pulse shape filtering.

The peak-to-average-power ratio (PAPR) increase caused by DDST training has been studied in [P3],[P4] and simple symbol level limiter has been proposed to reduce the PAPR increase. In addition, in [P4] the throughput performance with respect to E_b/N_0, of TDMT and DDST based systems was compared with different modulations and receiver antenna diversity orders. In [P5], the throughput of DDST and TDMT was compared in multiple-input multiple-output (MIMO) communications with various antenna configurations and modulations.

This thesis is organized such that the following Chapter 2 describes the basic principles of TDMT, DDST and SIT and defines the least squares (LS) and LMMSE channel esimators for each training method. First the channel estimators are defined for single-input single-output scheme (SISO), after which the MIMO scheme is briefly described. Then, in Chapter 3, a short introduction to channel equalization and a few words about discrete Fourier transform (DFT) and FB based equalizers and their differences are given. These chapters provide basic theory that is required as background for understanding all the publications.

In Chapter 4, a short introduction to iterative detection of data symbols with DDST is provided. With DDST, because the data symbols under detection suffer from the

interference caused by the missing cyclic mean component, the simplest solution is to rely on iterative processing in which we in turns estimate the data symbols and then the cyclic mean of these data symbols. This chapter provides background information for understanding receiver operation and performance results related to publications [P3]-[P5].

Chapter 5 provides an unpublished analysis on the DDST symbol error probability based on the cyclic mean error distribution with additive white Gaussian noise (AWGN) channel. In addition, an alternative approach based on mutual information between transmitted and received symbols is used to derive bit error probabilities for DDST with or without using a priori information in the receiver. This chapter provides unpublished analysis that supports the discussion related to cyclic mean error term distribution given in [P4].

In Chapter 6, the PAPR modeling is shortly described. This chapter supports studies presented in publications [P3],[P4]. Then, in Chapter 7, the contributions of this thesis are wrapped up and future research topics are provided.

1.4 *Author's Contribution to the Publications*

The research topics related to SIT and DDST on wide band SC transmission originated from a separate study on narrow band peer-to-peer SC communications using SIT [101]. In all of the publications, [P1]-[P5] the derivation of the main results and most of the writing effort were done by the author. In [P1], M.Sc. Jukka Talvitie provided guidance for SIT based systems. In [P3],[P4], M.Sc Jukka Talvitie provided support and several discussions on the PAPR and PAPR reduction related topics, helping to improve the quality of the publications. Prof. Markku Renfors has provided valuable comments on the content and final structure of all the publications [P1]-[P5].

2. CHANNEL ESTIMATION

Because the wireless medium causes destructive distortion on the transmitted signal, the channel estimation is a crucial part of any wireless receiver utilizing coherent detection. Channel estimation is typically based on known reference symbols, but also blind or semi blind algorithms exist that use none or minimal amount of reference symbols (in this thesis reference symbols, training symbols, and pilot symbols or signals are used as synonyms, referring to some known symbols in the transmitted sequence that we utilize to obtain channel estimates).

After time and frequency domain synchronization, channel estimation is the most important task to allow receiver to reliably detect the transmitted signal. The higher the required throughput is, mapping to higher modulation order and lower coding rate, the higher the channel estimation accuracy has to be. The reference symbol layout design based on knowledge of the worst case channel delay spreads and mobility affects the maximum accuracy that can be achieved from a certain training scheme. Also, power boosting of reference symbols may improve the system performance, at least in terms of the channel estimation mean squared error (MSE).

In this section, we will look at the least squares (LS) and linear minimum mean squared error (LMMSE) channel estimators for superimposed training (SIT), time-domain multiplexed training (TDMT) and data-dependent superimposed training (DDST) cases and evaluate their MSE performance. Comments on the assumptions for achieving the presented MSE values with SIT, TDMT and DDST are also given.

2.1 System Model

Both, LS and LMMSE estimation algorithms are well known and widely used in several different estimation tasks, including channel estimation in wireless receivers. An excellent book explaining these algorithms is given by Kay [62].

Before we go into defining the channel estimators, we have to define the problem setup and related variables. We start by defining TDMT and then SIT and DDST.

Discussion on TDMT and related channel estimation can be found for example in [89]. In [32, 33] both TDMT and SIT are discussed in a more analytical level and in [115,116] the authors concentrate on the achievable data rates with SIT using either sequential reception or joint channel and symbol estimation. Additional discussion on SIT can be found e.g. in [6, 18, 19, 23, 37–39, 47, 50, 51, 59, 65, 72, 74, 75, 77–80, 86–88, 106, 107, 110, 113, 125, 126] and references therein. Narrow band and low rate communications comparison between TDMT and SIT is provided in [101]. For DDST, the channel estimation aspect is discussed at least in [8, 45, 46, 68, 74–76, 81, 105] and references therein.

In our studies, we have assumed that each packet is protected by an transmission interval or guard period longer than the maximum channel delay. This would correspond to discontinuous transmission of packets. For continuous transmission cyclic prefix (CP) is preferred to avoid transmission gaps in the transmitted signal. Transmission intervals and CPs are used to prevent inter-packet interference (IPI), similar to OFDM. In this chapter, we derive the results assuming a CP to be used. The matrix used to extend the packet of length N by $L_c - 1$ samples is given as

$$\mathbf{M}_{AddCP}(N, L_c - 1) = \begin{bmatrix} \mathbf{0}_{L_c-1 \times N-L_c+1} & \mathbf{I}_{L_c-1} \\ & \mathbf{I}_N \end{bmatrix}. \qquad (2.1.1)$$

and the matrix used to remove the CP is defined as

$$\mathbf{M}_{RemoveCP}(N, L_c - 1) = \begin{bmatrix} \mathbf{0}_{N \times L_c-1} & \mathbf{I}_N \end{bmatrix}. \qquad (2.1.2)$$

Throughout this thesis we will assume that pilot and data symbols have unit variance, $\sigma_p^2 = 1$ and $\sigma_d^2 = 1$, respectively. Therefore, in the equations defining the transmitted signals, we include the power scaling factor γ defining the fraction of total power divided between training signal and data signal, where signal now refers to the power scaled symbols. In addition, for TDMT a scaling factor, λ, for pilot signal is required to force it to have fraction γ of the total power and is defined as

$$\lambda = \frac{\gamma N_d}{N_p + N_{cp}}, \qquad (2.1.3)$$

where N_d is the number of data symbols, N_p is the number of pilot symbols and N_{cp} is the number symbols in the cyclic prefix. This scaling factor contains power allocation

for the pilot based CP. For DDST and SIT, the CP is taken from the end of the sum signal, and therefore they require their own scaling factor for removing the CP power from the data signal. This scaling factor is defined as

$$\beta = \frac{N_d}{N_d + N_{cp}}. \tag{2.1.4}$$

We will assume that the channel length is always limited to maximum duration of L_c samples. Because the training signal is known by the receiver and thus does not carry any information, we want to minimize the overhead it causes. This leads to selecting the length of total training sequence with TDMT, $\mathbf{p}_{TDMT,cp}$, to be equal to $2L_c - 1$. This value contains now CP of length $N_{cp} = L_c - 1$ and actual pilot vector of length $N_p = L_c$. Defining the transmitted pilot sequence with CP as $\mathbf{p}_{TDMT,cp} = \sqrt{\lambda} \mathbf{M}_{AddCP}(N_p, L_c - 1)\mathbf{p}_{TDMT} = \sqrt{\lambda}[p(1)\ p(2)\ \dots\ p(L_c - 2)\ p(L_c - 1)\ p(0)\ p(1)\ \dots\ p(L_c - 2)\ p(L_c - 1)]^T$, we end up with circular matrix for efficient channel estimation in the receiver.

For TDMT the transmitted signal $\mathbf{x}_{TDMT}$ is defined as

$$\mathbf{x}_{TDMT,cp} = \begin{bmatrix} \sqrt{\lambda} \mathbf{M}_{AddCP}(N_p, L_c - 1)\mathbf{p}_{TDMT} \\ \sqrt{(1 - \gamma)}\mathbf{d}_{TDMT} \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{p}_{TDMT,cp} \\ \sqrt{(1 - \gamma)}\mathbf{d}_{TDMT} \end{bmatrix}, \tag{2.1.5}$$

where $\mathbf{p}_{TDMT,cp}$ is the training sequence extended with CP and $\mathbf{d}_{TDMT}$ is the data symbol vector.

When compared to TDMT, SIT and DDST follow completely different ideology. In TDMT the data and training symbols are multiplexed in time, whereas in SIT and DDST we add the scaled training and data symbols together. SIT is defined as

$$\mathbf{x}_{SIT} = \sqrt{\gamma}\mathbf{p}_{c,SIT} + \sqrt{1 - \gamma}\mathbf{d}_{SIT}$$
$$= \sqrt{\gamma}(\mathbf{I}_{N_c \times 1} \otimes \mathbf{p}_{SIT}) + \sqrt{1 - \gamma}\mathbf{d}_{SIT}, \text{ and} \tag{2.1.6}$$

$$\mathbf{x}_{SIT,cp} = \sqrt{\beta}\mathbf{M}_{AddCP}(N_d, L_c - 1)\mathbf{x}_{SIT} \tag{2.1.7}$$

where $\mathbf{p}_{c,SIT}$ is the cyclic pilot sequence consisting of N_c copies of the basis pilot vector $\mathbf{p}_{SIT}$ and $\mathbf{x}_{SIT,cp}$ is the sum signal extended with CP.

For DDST, first we define a matrix $\mathbf{J}(a,b,c)$ that is used to evaluate a copies of the cyclic mean over b cycles of length c as

$$\mathbf{J}(a,b,c) = \frac{1}{b}\mathbf{1}_{a\times b}\otimes\mathbf{I}_c, \tag{2.1.8}$$

where $\otimes$ defines the Kronecker product and $\mathbf{1}_{a\times b}$ is a matrix of ones with a rows and b columns. Now, the transmitted signal is defined as

$$
\begin{aligned}
\mathbf{x}_{DDST} &= \sqrt{\gamma}\,\mathbf{p}_{c,DDST} + \sqrt{1-\gamma}\,\tilde{\mathbf{d}}_{DDST} \\
&= \sqrt{\gamma}\,(\mathbf{1}_{N_c\times 1}\otimes\mathbf{p}_{DDST}) + \sqrt{1-\gamma}\,\sqrt{\frac{N_c}{N_c-1}}(\mathbf{d}_{DDST}+\mathbf{p}_d) \\
&= \sqrt{\gamma}\,(\mathbf{1}_{N_c\times 1}\otimes\mathbf{p}_{DDST}) + \sqrt{1-\gamma}\,\sqrt{\frac{N_c}{N_c-1}}(\mathbf{I}_N - \mathbf{J}(N_c,N_c,L_c))\mathbf{d}_{DDST},\text{ and}
\end{aligned}
\tag{2.1.9}
$$

$$\mathbf{x}_{DDST,cp} = \sqrt{\beta}\mathbf{M}_{AddCP}(N_d, L_c - 1)\mathbf{x}_{DDST}, \tag{2.1.10}$$

where $\mathbf{p}_{c,DDST}$ is the cyclic pilot sequence consisting of N_c copies of the basis pilot vector $\mathbf{p}_{DDST}$. Here $\mathbf{p}_d = -\mathbf{J}(N_c,N_c,L_c)\mathbf{d}_{DDST}$ is now the self interference term, defined as the negative cyclic mean component of the data vector $\mathbf{d}_{DDST}$. In (2.1.9), the power normalization term $\sqrt{N_c/(N_c-1)}$ for the distorted data sequence is used to return the power of the data component to unity after removing the cyclic mean for fair comparison. From (2.1.9), we can see that in the case of DDST we can talk about data dependent pilot signal, if we consider the effective pilot to be $\mathbf{p}_{eff} = \mathbf{p}_{c,DDST} + \mathbf{p}_d$, or we can describe the transmitted signal as data dependent (or time variant) modulation with SI training where the data dependent modulation is defined as $\mathbf{d}_{timevar} = (\mathbf{I}_N - \mathbf{J}(N_c,N_c,L_c))\mathbf{d}_{DDST}$. Here the modulation of symbols changes inside one cycle and per packet basis. In the receiver, in the channel estimation we have a pure SI training signal without self interference and in the symbol detection we have the time variant modulation component that is generated by the removal of the cyclic mean of the data vector. Because of this per packet time varying modulation, we may use iterative symbol detectors to improve the symbol detection probability, as we will see in Section 4.

Also, with DDST transmission the overhead to achieve circular pilot matrix without inter-packet interference (IPI) is equal to $2L_c - 1$ lost dimensions (see comments

on equality of lost dimensions between TDMT and DDST from [17]). If we want to discard the CP as done in [76], most of the error comes from the tailing of the previous packet. Also, self interference is caused by missing part of the samples needed to perfectly cancel the interference from our own data transmission. In [76] it was claimed that increase in the interference is not significant, but this depends on the used modulation and coding scheme (MCS).

If the CP is not used between packets, we propose that also the samples belonging to the first cycle should be ignored in the channel estimation process, in order to remove the IPI. This will cause increase in the self interference term, but if the number of copies N_c is relatively large, then the self interference has minor effect on the performance. Also, because we assume that higher fraction of power is on the data symbols, typically the increase of the self interference is smaller than the error increase caused by previous packet's data symbols.

We have only studied discontinuous packet transmission assuming that the channel is empty before transmitting a packet and after the transmission we wait a time period longer than the channel coherence time to get independent transmissions. We have not been interested in continuous time performance because we first want to understand the performance limits of DDST and especially the relative performance with respect to TDMT. We emphasize that in continuous transmission through fading channel the tracking properties of SIT and DDST may exceed TDMT because of the training signal is always available, but the verification of this assumption requires further studies. Also, the performance of DDST and SIT in the case of impulse like interference could allow significantly better channel estimation and interference cancellation than TDMT.

2.2   Least Squares Solution

The least squares (LS) solution is well known and typically is the first thing to try when estimating parameters from a linear set of equations. In the basic channel estimation problems the least squares solution is also the maximum likelihood solution. This can be easily shown, e.g. for TDMT.

Let us model the received signal with TDMT as $\mathbf{y}_{TDMT} = \mathbf{H}\mathbf{x}_{TDMT,cp} + \mathbf{n}$, where $\mathbf{n}$ is complex, white Gaussian noise with variance σ_n^2. Let us now separate the pilot and data portion of the received vector as

$$\mathbf{y}_{TDMT} = \begin{bmatrix} \mathbf{y}_{p,cp} \\ \mathbf{y}_d \end{bmatrix} = \begin{bmatrix} \mathbf{H}_p \\ \mathbf{H}_d \end{bmatrix} \begin{bmatrix} \mathbf{p}_{TDMT,cp} \\ \sqrt{1-\gamma}\mathbf{d}_{TDMT} \end{bmatrix} + \begin{bmatrix} \mathbf{n}_{p,cp} \\ \mathbf{n}_d \end{bmatrix}, \quad (2.2.1)$$

from which we are interested only in the pilot related part of the signal $\mathbf{y}_{p,cp}$. Ignoring now the $L_c - 1$ first samples generating the CP and scaling the received power, we end up with received training vector given as

$$\begin{aligned} \mathbf{y}_p &= \sqrt{\lambda^{-1}}\mathbf{M}_{RemoveCP}\mathbf{y}_{p,cp} \\ &= \mathbf{P}_{TDMT}\mathbf{h} + \sqrt{\lambda^{-1}}\mathbf{n}_p. \end{aligned} \quad (2.2.2)$$

The complex noise is defined as $\mathbf{n}_p \in \mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I}_{L_c})$ and $\mathbf{P}_{TDMT}$ is a circular matrix with first column defined by the basis pilot vector $\mathbf{p}_{TDMT}$. Now, the maximum likelihood estimate (ML) is defined as

$$\begin{aligned} \hat{\mathbf{h}}_{ML,TDMT} &= \arg_{\mathbf{h}} \max f(\mathbf{y}_p|\mathbf{h}) \\ &= \cdots = \arg_{\mathbf{h}} \max \|\mathbf{y}_p - \mathbf{P}\mathbf{h}\|^2. \end{aligned} \quad (2.2.3)$$

Next, because the ML channel estimation problem is a convex problem, taking the derivative with respect to the channel vector $\mathbf{h}$ and by setting it to zero, we end up with ML channel estimate given as

$$\hat{\mathbf{h}}_{ML,TDMT} = (\mathbf{P}_{TDMT}^H \mathbf{P}_{TDMT})^{-1} \mathbf{P}_{TDMT}^H \mathbf{y}_p = \hat{\mathbf{h}}_{LS,TDMT}, \quad (2.2.4)$$

and we notice that the ML estimate is equal to the LS channel estimate. The well known mean squared error (MSE) for an estimator like this is given as

$$\begin{aligned} MSE(\hat{\mathbf{h}}_{LS,TDMT}) &= \mathrm{E}(\|\mathbf{h} - \hat{\mathbf{h}}\|^2) = \mathrm{trace}\{\mathrm{E}((\mathbf{h} - \hat{\mathbf{h}})(\mathbf{h} - \hat{\mathbf{h}})^H)\} \\ &= \cdots = \lambda^{-1}\sigma_n^2 \mathrm{trace}\{(\mathbf{P}^H\mathbf{P})^{-1}\}. \end{aligned} \quad (2.2.5)$$

With LS channel estimation there is a direct connection between the noise variance and the variance of the pilot vector, especially if Chadoff-Chu sequences are used [27]. These sequences have ideal impulse like circular correlation properties and we

end up with $MSE(\hat{\mathbf{h}}_{LS,TDMT}) = \sigma_n^2/(\lambda\sigma_p^2)$. For DDST, Chadoff-Chu like sequences were proposed in [87], where they were referred to as chirp sequences. It was shown that they are optimal in a sense that by using them we end up with unitary circulant matrices that are not dependent of the channel in the SIT scenario. We have also used Chadoff-Chu sequences with DDST in all of the publications, because of the unitary circulant matrix property.

The received signal with SIT after removing the CP is defined as

$$
\begin{aligned}
\mathbf{y}_{SIT} &= \sqrt{\beta^{-1}}\mathbf{M}_{RemoveCP}\mathbf{y}_{SIT,cp} \\
&= \mathbf{H}\mathbf{x}_{SIT} + \sqrt{\beta^{-1}}\mathbf{n} \\
&= \mathbf{X}_{SIT}\mathbf{h} + \sqrt{\beta^{-1}}\mathbf{n}
\end{aligned}
\tag{2.2.6}
$$

where $\mathbf{H}$ is the full channel matrix and $\mathbf{X}_{SIT}$ is the full data matrix with dimensions $(N_d \times L_c)$. The cyclic mean of the received signal is given as

$$
\begin{aligned}
\mathbf{y}_{M,SIT} &= \tfrac{1}{\sqrt{\gamma}}\mathbf{J}(1,N_c,L_c)\mathbf{y}_{SIT} \\
&= \tfrac{1}{\sqrt{\gamma}}(\mathbf{X}_{M,SIT}\mathbf{h} + \sqrt{\beta^{-1}}\mathbf{n}_M) \\
&= \mathbf{P}_{SIT}\mathbf{h} + \tfrac{1}{\sqrt{\gamma}}(\sqrt{1-\gamma}\mathbf{D}_M\mathbf{h} + \sqrt{\beta^{-1}}\mathbf{n}_M),
\end{aligned}
\tag{2.2.7}
$$

where $\mathbf{P}_{SIT}$ is a circular matrix with the basis pilot vector $\mathbf{p}_{SIT}$ as the first column and $\mathbf{D}_M$ is the circular matrix containing the cyclic mean values of the data vector $\mathbf{d}$. The LS estimator solution for SIT is defined as

$$
\begin{aligned}
\hat{\mathbf{h}}_{LS,SIT} &= (\mathbf{P}_{SIT}^H\mathbf{P}_{SIT})^{-1}\mathbf{P}_{SIT}^H\mathbf{y}_{M,SIT} \\
&= \mathbf{h} + (\mathbf{P}_{SIT}^H\mathbf{P}_{SIT})^{-1}\mathbf{P}_{SIT}^H[\tfrac{\sqrt{1-\gamma}}{\sqrt{\gamma}}\mathbf{D}_M\mathbf{h} + \tfrac{1}{\sqrt{\beta\gamma}}\mathbf{n}_M].
\end{aligned}
\tag{2.2.8}
$$

The noise variance after the cyclic mean operation corresponds to $\sigma_{n_M}^2 = \sigma_n^2/(\beta N_c)$ and the variance of the self-interference term is $\sigma_{d_M}^2 = \sigma_d^2/N_c$. For SIT the derivation of the MSE for the LS channel estimator is more difficult, because the self interference term $\mathbf{D}_M\mathbf{h}$ depends on the channel distribution. For comparison purposes, we can easily evaluate the MSE performance in a unity channel, for which the channel response is $\mathbf{h}_{\mathbf{unity}} = [1\ \mathbf{0}_{1\times L_c-1}]^T$. The MSE for the LS channel estimator with SIT in

a unity channel is given as

$$
\begin{aligned}
MSE(\hat{\mathbf{h}}_{LS,SIT}) \quad &= \text{trace}\{\mathrm{E}[(\mathbf{h}_{unity} - \hat{\mathbf{h}}_{unity})(\mathbf{h}_{unity} - \hat{\mathbf{h}}_{unity})^H]\} \\
&= \frac{(1-\gamma)\sigma_d^2 + \beta^{-1}\sigma_n^2}{\gamma N_c}\text{trace}\{(\mathbf{P}^H\mathbf{P})^{-1}\}.
\end{aligned}
\tag{2.2.9}
$$

With DDST, based on (2.1.9), we can define the received signal after discarding the CP as

$$
\begin{aligned}
\mathbf{y}_{DDST} \quad &= \sqrt{\beta^{-1}}\mathbf{M}_{RemoveCP}\mathbf{y}_{DDST,cp} \\
&= \mathbf{H}\mathbf{x}_{DDST} + \sqrt{\beta^{-1}}\mathbf{n} \\
&= \mathbf{X}_{DDST}\mathbf{h} + \sqrt{\beta^{-1}}\mathbf{n},
\end{aligned}
\tag{2.2.10}
$$

where $\mathbf{H}$ is the full channel matrix and $\mathbf{X}_{DDST}$ is the full data matrix with dimensions $(N \times L_c)$. The cyclic mean of the received signal is given as

$$
\begin{aligned}
\mathbf{y}_{M,DDST} \quad &= \frac{1}{\sqrt{\gamma}}\mathbf{J}(1,N_c,L_c)\mathbf{y}_{DDST} \\
&= \frac{1}{\sqrt{\gamma}}(\mathbf{X}_{M,DDST}\mathbf{h} + \sqrt{\beta^{-1}}\mathbf{n}_M) \\
&= \mathbf{P}_{DDST}\mathbf{h} + \frac{1}{\sqrt{\beta\gamma}}\mathbf{n}_M,
\end{aligned}
\tag{2.2.11}
$$

where $\mathbf{P}_{DDST}$ is a circulant matrix with the basis pilot vector $\mathbf{p}_{DDST}$ as the first column. The LS solution for DDST is then given as

$$
\begin{aligned}
\hat{\mathbf{h}}_{LS,DDST} \quad &= \frac{1}{\sqrt{\gamma}}(\mathbf{P}_{DDST}^H\mathbf{P}_{DDST})^{-1}\mathbf{P}_{DDST}^H\mathbf{y}_{M,DDST} \\
&= \mathbf{h} + \frac{1}{\sqrt{\beta\gamma}}(\mathbf{P}_{DDST}^H\mathbf{P}_{DDST})^{-1}\mathbf{P}_{DDST}^H\mathbf{n}_M.
\end{aligned}
\tag{2.2.12}
$$

Notice that the terms related to data vector, $\mathbf{d}_{DDST}$, disappear by taking the cyclic mean over the received vector. This is the main motivator for using DDST instead of more traditional SI. In SI training, the initial self interference from the data sequence can be so significant that obtaining reliable channel estimate is difficult. Through iterative processing the quality of the channel estimate can be improved if the accuracy of the initial estimate is above certain threshold. With DDST the channel estimate is always good (assuming good synchronization, see [6, 8] for details). The iterative processing is then required for finding the missing cyclic mean of the data vector.

We believe that this is simpler task, because the data vector is made of well known discrete symbols and finding the missing part should be simpler than finding a continuous channel estimate. See Appendix A for a simple explanation how DDST can be looked as a frequency domain multiplexed (FDM) training with SC transmission and why it destroys information bearing subcarriers in OFDM transmission.

The MSE for the LS channel estimator with DDST is given as

$$
\begin{aligned}
MSE(\hat{\mathbf{h}}_{LS,DDST}) \quad &= \text{trace}\{\mathrm{E}[(\mathbf{h}-\hat{\mathbf{h}})(\mathbf{h}-\hat{\mathbf{h}})^H]\} \\
&= \frac{\sigma_n^2}{\beta\gamma N_c}\text{trace}\{(\mathbf{P}^H\mathbf{P})^{-1}\}.
\end{aligned}
\tag{2.2.13}
$$

If we assume that TDMT and DDST use the same basis training vector $\mathbf{p}$ for channel estimation, allocate the same fraction of total data power, γ, to pilots ,use the same cyclic prefix length, N_{cp}, and have equal data vector length, N_d, we can define the ratio between DDST and TDMT LS estimator MSEs as

$$
\begin{aligned}
\frac{MSE(\hat{\mathbf{h}}_{LS,DDST})}{MSE(\hat{\mathbf{h}}_{LS,TDMT})} \quad &= \frac{\frac{\sigma_n^2}{\beta\gamma N_c}\text{trace}\{(\mathbf{P}^H\mathbf{P})^{-1}\}}{\frac{\sigma_n^2}{\lambda}\text{trace}\{(\mathbf{P}^H\mathbf{P})^{-1}\}} = \frac{\lambda}{\beta\gamma N_c} \\
&= \frac{\frac{\gamma N_d}{N_p+N_{cp}}}{\frac{N_d\gamma N_c}{N_d+N_{cp}}} = \frac{N_d+N_{cp}}{N_c(N_p+N_{cp})},
\end{aligned}
\tag{2.2.14}
$$

and if we assume that $N_c = N_d/N_p$, so that the length of the cyclic pilot sequence is exactly the length of the data sequence, we end up with result

$$
\frac{MSE(\hat{\mathbf{h}}_{LS,DDST})}{MSE(\hat{\mathbf{h}}_{LS,TDMT})} \quad = \frac{N_dN_p+N_pN_{cp}}{N_dN_p+N_dN_{cp})} = \frac{1+\frac{N_{cp}}{N_d}}{1+\frac{N_{cp}}{N_p}}.
\tag{2.2.15}
$$

We notice that always when DDST has more than one copy of pilot sequence on top of data sequence ($N_d > N_p$), the channel estimation MSE with DDST is smaller than with TDMT, even though both methods allocate same fraction of total power, γ, for training signal. Interesting fact is also that the ratio of MSE performance is independent of the actual value of γ, because it is not present in the final form.

In Fig. 1 a), example of the MSE ratio as a function of pilot vector length and data vector length is provided. It should be noted that not all values for N_p and N_d are valid. As discussed already in [17], if $N_p \leq N_d < 2N_p$, there is only one copy of the basis pilot sequence on top of the data sequence. Then, the "cyclic"
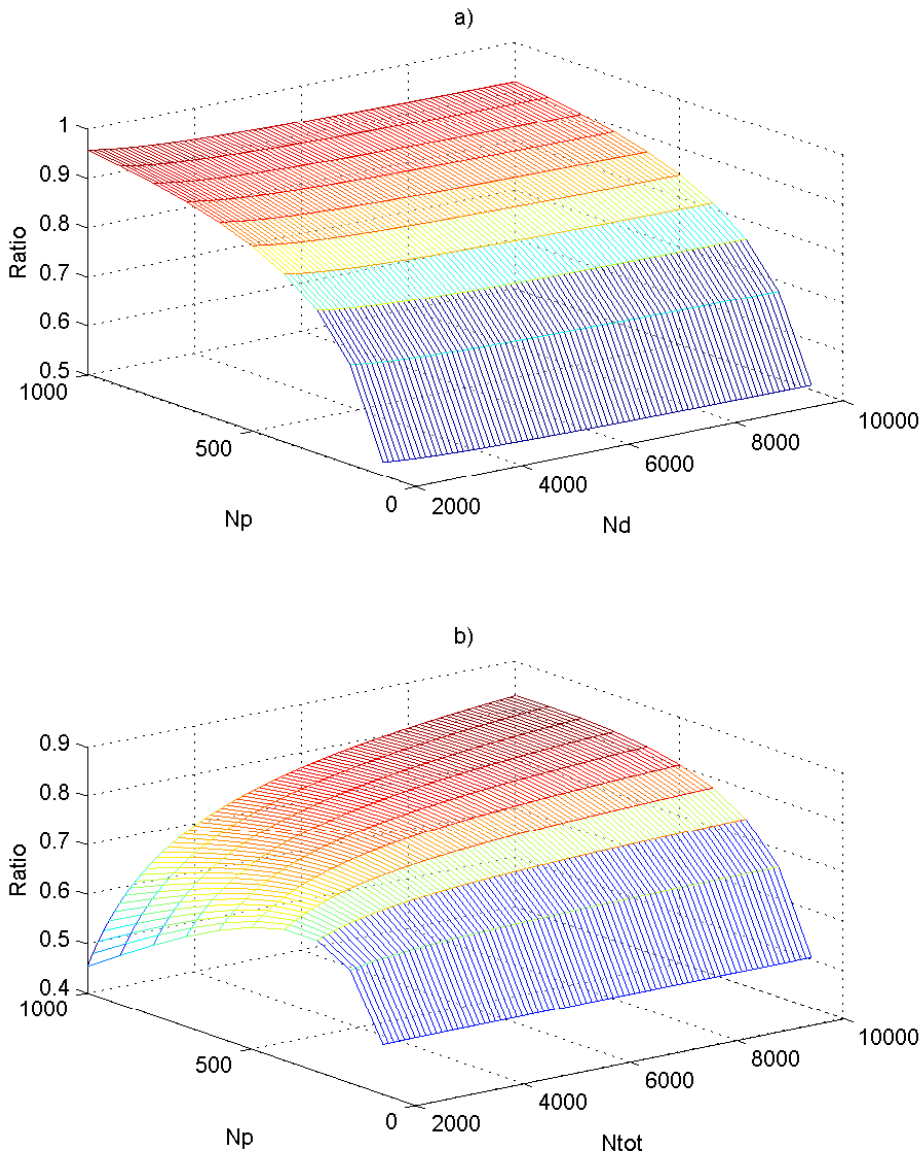
Fig. 1. *Comparison between LS based channel estimation MSE between DDST and TDMT. In a), the length of the data vector, N_d, is equal for both systems. In b), the total length of the packet, N_{tot}, including cyclic prefix, pilot sequence and data sequence, is equal for both systems.*

mean corresponds to the data sequence itself. Now, to remove the self interference, we would have to remove the data sequence itself. This means that we would not transmit any information on these symbols and transmission would be pointless. In theory, DDST works if $N_d \geq 2N_p$, but with very small number of copies the self interference term causes significant degradation in the detection process.

Another interesting aspect for comparing the channel estimation MSEs between DDST and TDMT is obtained, when we set the total packet length, N_{tot} to be equal for both systems. This indicates that $N_{d,DDST} = N_{p,TDMT} + N_{d,TDMT}$, assuming that both methods use the same cyclic prefix length N_{cp}. This comparison is more valid from the real system implementation point of view, because independent of the training method, typically the packet duration is fixed to discrete values by the system design. The MSE ratio between these two systems, assuming the same N_{tot}, is given as

$$\frac{MSE(\hat{\mathbf{h}}_{LS,DDST})}{MSE(\hat{\mathbf{h}}_{LS,TDMT})} = \frac{N_{tot}N_p(N_{tot} - N_{cp} - N_p)}{(N_{tot} - N_{cp})^2(N_p + N_{cp})}. \tag{2.2.16}$$

Example of the ratio of the channel estimation MSEs, assuming same total packet length, is shown in Fig. 1 b). We can notice how the total packet length limitation improves the relative channel estimation performance of DDST when the number of pilot symbols is relatively high with respect to the number of data symbols.

2.3   Linear Minimum Mean Squared Error Solution

After looking at LS channel estimators, an obvious extension is to consider LMMSE channel estimators. The used LMMSE estimators are based on [62]. The LMMSE estimator is designed to minimize the expected squared error between the true channel and the channel estimate, defined as $\hat{\mathbf{h}} = \min_{\hat{\mathbf{h}}} E\left[(\mathbf{h} - \hat{\mathbf{h}})^2\right]$. Given that we have a linear distortion model

$$\mathbf{y} = \mathbf{P}\mathbf{h} + \mathbf{n}, \tag{2.3.1}$$

then the LMMSE estimator is given as

$$\begin{aligned}
\hat{\mathbf{h}}_{LMMSE} &= E(\mathbf{h}) + \mathbf{C}_{hy}\mathbf{C}_{yy}^{-1}(\mathbf{y} - E(\mathbf{y})) \\
&= E(\mathbf{h}) + \mathbf{C}_{hh}\mathbf{P}^H(\mathbf{P}\mathbf{C}_{hh}\mathbf{P}^H + \mathbf{C}_n)^{-1}(\mathbf{y} - E(\mathbf{y})) \\
&= E(\mathbf{h}) + (\mathbf{C}_{hh}^{-1} + \mathbf{P}^H\mathbf{C}_n^{-1}\mathbf{P})\mathbf{P}^H\mathbf{C}_n^{-1}(\mathbf{y} - E(\mathbf{y})),
\end{aligned} \tag{2.3.2}$$

and the estimation error, $\varepsilon = \mathbf{h} - \hat{\mathbf{h}}$, has a covariance matrix defined as

$$
\begin{aligned}
\mathbf{C}_\varepsilon &= E_{y,h}(\varepsilon\varepsilon^H) \\
&= \mathbf{C}_{hh} - \mathbf{C}_{hh}\mathbf{P}^H(\mathbf{P}\mathbf{C}_{hh}\mathbf{P}^H + \mathbf{C}_n^{-1})^{-1}\mathbf{P}\mathbf{C}_{hh} \\
&= (\mathbf{C}_{hh}^{-1} + \mathbf{P}^H\mathbf{C}_n^{-1}\mathbf{P})^{-1},
\end{aligned}
\tag{2.3.3}
$$

where the diagonals give the minimum MSE per sample. In the results presented in this thesis, we compare the sum MSE performance, which relates to the error covariance matrix as

$$
MSE(\hat{\mathbf{h}}_{x,LMMSE}) = \text{trace}\{\mathbf{C}_\varepsilon\}.
\tag{2.3.4}
$$

2.3.1 LS-LMMSE Approximative Channel Estimator

One relevant and somewhat open question is that how should the channel covariance matrix $\mathbf{C}_h = E(\mathbf{h}\mathbf{h}^H)$ be estimated in the receiver. We have chosen to approximate the covariance matrix by a diagonal matrix obtained from the power response of the LS channel estimate. Other simple solutions would be to calculate instantaneous covariance matrix estimate based on the sample covariance. In a continuous time channel, in which we operate in the real world, by averaging sample covariance matrices over certain time window or by using exponential weighting (time memory) we may achieve improved estimates of the channel covariance. In our model the channel covariance matrix estimate is given as

$$
\hat{\mathbf{C}}_{\hat{h},x} = \text{diag}([|\hat{h}_{LS,x}(0)|^2 \ |\hat{h}_{LS,x}(1)|^2 \ \cdots \ |\hat{h}_{LS,x}(L_c-1)|^2]^T),
\tag{2.3.5}
$$

where subindex x indicates the used training method, and may be DDST, SIT, or TDMT. This approximated channel coefficient covariance matrix contains the required prior information of the channel that is needed for the LS-LMMSE channel estimator (also noted as ML-LMMSE channel estimator in the publications [P2] and [P3]) used in publications [P2]-[P5].

Furthermore, we always assume that the data sequence, channel response and independently and identically distributed (i.i.d.) noise are zero mean random variables.

Table 3. *Parameter definitions for unified presentation for the LS-LMMSE channel estimator.
Here, $\sigma_{w,x}^2$ represents the training dependent interference variance.*

x	TDMT	SIT	DDST
$\sigma_{w,x}^2$	σ_n^2/λ	$[(1-\gamma)\sigma_d^2 + \sigma_n^2/\beta]/(\gamma N_c)$	$\sigma_n^2/(\beta\gamma N_c)$
$\mathbf{z}$	$\mathbf{y}_p$	$\mathbf{y}_{M,SIT}$	$\mathbf{y}_{M,DDST}$

Now, for all of the used channel estimation methods, a unified LS-LMMSE channel estimator is given as

$$\hat{\mathbf{h}}_{x,LS-LMMSE} = (\sigma_{w,x}^2 \hat{\mathbf{C}}_{\hat{h},x}^{-1} + \mathbf{P}^H\mathbf{P})^{-1}\mathbf{P}^H\mathbf{z}, \qquad (2.3.6)$$

where x indicates the used training method, and may be DDST, SIT, or TDMT and $\mathbf{z}$ is the corresponding received sequence for the used channel estimation method. The approximated MSE for the LS-LMMSE channel estimator is given as

$$MSE(\hat{\mathbf{h}}_{x,LS-LMMSE}) = \cdots = \text{trace}\{(\mathbf{C}_h^{-1} + \sigma_{w,x}^{-2}\mathbf{P}^H\mathbf{P})^{-1}\}. \qquad (2.3.7)$$

Depending on the used training method, variables $\mathbf{z}$ and $\sigma_{w,x}^2$ have values as defined in Table 3.

2.4   MIMO Channel Estimation

Multiple-input multiple-output (MIMO) communications are extension to the single-input single output (SISO) communications discussed in the previous section. The method we utilize for MIMO channel estimation relies on the Zadoff-Chu sequences [27], also known as constant amplitude zero autocorrelation (CAZAC) waveforms [1] or optimal channel independent (OCI) pilot sequence [87]. The Zadoff-Chu sequence of length N_p is defined as

$$p_s(n) = e^{-j\frac{\pi sn(n+1+2\eta)}{N_p}}, \qquad (2.4.1)$$

where sample index n is limited by $0 \leq n < N_p - 1$, sequence index s is limited by $0 < s < N_p$ and $\eta \in \mathbb{Z}$. We have only considered case where $\eta = 0$ and N_p is a prime. The sequence length N_p does not have to be prime, but using a non-prime length

significantly reduces the number of good sequences available for a certain length. With N_p being prime we get $N_p - 2$ sequences with desired correlation properties.

The main beauty of Zadoff-Chu sequences is that assuming the length of the sequence N_p to be prime and having same η value, the circular autocorrelation is zero with all non-zero cyclic shifts and the cyclic cross correlation between two sequences results in constant $1/\sqrt{N_p}$. Also, they have a constant amplitude response in the frequency domain which is required for a good channel estimation sequence.

When choosing a Zadoff-Chu sequence for MIMO transmission, the length of the used sequence has to be $N_p \geq N_{Tx}L_c$ in order for us to estimate all the channels per Rx antennae. We assume that we either know L_c or we have an estimate or system design constrain for the maximum allowed channel length. We have used the knowledge of the true channel length when defining the minimum length for the pilot sequence.

We have also studied the effects of truncating the channel estimator with DDST to be shorter than the expected maximum channel delay. We have noticed, as expected, that there is a trade off between the channel estimation accuracy loss and the gain from increasing the number of cyclic copies in the case of DDST. Therefore, especially in noisy environments where robust modulation and coding schemes (MCS) are used, it might be useful to reduce the channel estimation accuracy to obtain better noise reduction performance.

In addition to choosing the right prime number for the pilot length, careful placing of the pilot symbols per data stream per-Tx-antenna has to be ensured [P5]. The channel estimation of per-Tx-antenna channels is based on correct cyclic shifts of the per-Tx-antenna pilot signal and correctly defined cyclic prefix to obtain full, circular pilot matrix in the receiver.

First of all, as in SISO case, we need a CP of length $L_c - 1$ samples to prevent inter-packet-interference (IPI). Now, this CP is taken per-Tx-antenna or stream (we assume that we always have as many Tx antennas as we have Tx spatial streams, we have not considered different space-time or space-frequency block codes in our studies) in order to achieve a full circular matrix presentation of the received signal. For TDMT, the indexing sets *IDX* for pilot sequence and pilot cyclic prefix per Tx antenna i are given as

$$IDX_{pilot,i} = \mod(k, N_p), \text{ where } (-i-1)L_c \leq k \leq -(i-1)L_c + N_p - 1, \quad (2.4.2)$$

and

$$IDX_{pilot,i,CP} = \text{mod}(k, N_p), \text{ where } -(i-1)L_c - L_c \leq k \leq -(i-1)L_c - 1, \quad (2.4.3)$$

respectively. With DDST, the indexing of a basis pilot vector per Tx antenna i is done based on (2.4.2), where as the CP is obtained by copying the $L_c - 1$ samples from the end of the packet, using the matrix defined in (2.1.1).

As an example, let us now consider of transmitting TDMT signal from 2 Tx antennas to 1 Rx antenna. Let the channel length be $L_c = 3$ and training signal length $N_p = 7$. Assuming that we transmit in to an empty channel (first packet, no distortion from previous transmission), we can rewrite the received signal in a matrix notation as

$$\mathbf{y}_{p,example\ with\ CP} = \mathbf{P}_{partial\ with\ CP}\mathbf{h}_{stacked} + \mathbf{n}_p$$

$$= [\mathbf{P}_{partial\ with\ CP,1} \ \mathbf{P}_{partial\ with\ CP,2} \ \mathbf{0}] \begin{bmatrix} \mathbf{h}_{1,1} \\ \mathbf{h}_{1,2} \\ \mathbf{0} \end{bmatrix} + \mathbf{n}_p$$

$$= \begin{bmatrix} p(5) & 0 & 0 & p(2) & 0 & 0 & 0 \\ p(6) & p(5) & 0 & p(3) & p(2) & 0 & 0 \\ p(0) & p(6) & p(5) & p(4) & p(3) & p(2) & 0 \\ p(1) & p(0) & p(6) & p(5) & p(4) & p(3) & 0 \\ p(2) & p(1) & p(0) & p(6) & p(5) & p(4) & 0 \\ p(3) & p(2) & p(1) & p(0) & p(6) & p(5) & 0 \\ p(4) & p(3) & p(2) & p(1) & p(0) & p(6) & 0 \\ p(5) & p(4) & p(3) & p(2) & p(1) & p(0) & 0 \\ p(6) & p(5) & p(4) & p(3) & p(2) & p(1) & 0 \end{bmatrix} \begin{bmatrix} h_{1,1}(0) \\ h_{1,1}(1) \\ h_{1,1}(2) \\ h_{1,2}(0) \\ h_{1,2}(1) \\ h_{1,2}(2) \\ 0 \end{bmatrix} + \begin{bmatrix} n_p(0) \\ n_p(1) \\ n_p(2) \\ n_p(3) \\ n_p(4) \\ n_p(5) \\ n_p(6) \\ n_p(7) \\ n_p(8) \end{bmatrix}.$$

$$(2.4.4)$$

If we now drop out the CP (2 first samples of $y_{p,example\ with\ CP}$), we end up with a partial, circular matrix. We can replace the partial, circular pilot matrix with a full, cir-

cular pilot matrix without affecting the outcome, because of the zeros in the stacked channel matrix. Then multiplying this result with the pseudo inverse of the full circular pilot matrix, $(\mathbf{P}^H\mathbf{P})^{-1}\mathbf{P}^H$, we end up with LS channel estimate defined as

$$
\hat{\mathbf{h}}_{stacked} =
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
h_{1,1}(0) \\
h_{1,1}(1) \\
h_{1,1}(2) \\
h_{1,2}(0) \\
h_{1,2}(1) \\
h_{1,2}(2) \\
0
\end{bmatrix}
+ (\mathbf{P}^H\mathbf{P})^{-1}\mathbf{P}^H
\begin{bmatrix}
n_p(2) \\
n_p(3) \\
n_p(4) \\
n_p(5) \\
n_p(6) \\
n_p(7) \\
n_p(8)
\end{bmatrix}.
$$

$$(2.4.5)$$

From this presentation, we can collect channels representing a certain Tx-Rx antenna pair. This example shows in principle how the CP per-Tx-antenna has to be chosen to obtain a full pilot matrix in the receiver. Then, from the properties of the used Zadoff-Chu sequences, we end up with stacked vector presentation of all the channels between all transmit antennae and studied receiving antenna.

The general notation for TDMT pilot signals received on antenna index r is given as

$$
\begin{aligned}
\mathbf{y}_{TDMT,p,r,CP} &= \sqrt{\tfrac{\lambda}{N_{Tx}}}\mathbf{P}_{CP}\mathbf{h}_{r,stacked} + \mathbf{n}_{p,r} \\
&= \sqrt{\tfrac{\lambda}{N_{Tx}}}\mathbf{P}_{CP}[\mathbf{h}_{r,1}^T, \mathbf{h}_{r,2}^T, \cdots, \mathbf{h}_{r,N_{Tx}}^T, \mathbf{0}^T]^T + \mathbf{n}_{r,p},
\end{aligned}
$$

$$(2.4.6)$$

where dividing with $\sqrt{N_{Tx}}$ is used to normalize the total transmitted power to unity. After removing the CP of the received signal and normalizing the power with $\sqrt{N_{Tx}/\lambda}$ the corresponding LS channel estimate is given as

$$
\begin{aligned}
\hat{\mathbf{h}}_{LS,TDMT,r} &= (\mathbf{P}_{TDMT}^H\mathbf{P}_{TDMT})^{-1}\mathbf{P}_{TDMT}^H\mathbf{y}_{TDMT,p,r} \\
&= \mathbf{h}_{r,stacked} + (\mathbf{P}_{TDMT}^H\mathbf{P}_{TDMT})^{-1}\mathbf{P}_{TDMT}^H\sqrt{N_{Tx}/\lambda}\,\mathbf{n}_{r,p}
\end{aligned}
$$

$$(2.4.7)$$

and the corresponding MSE over stacked channel vector is

$$\text{MSE}(\hat{\mathbf{h}}_{LS,TDMT,r}) = \frac{N_{Tx}\sigma_n^2}{\lambda}\text{trace}\{(\mathbf{P}_{TDMT}^H\mathbf{P}_{TDMT})^{-1}\}. \qquad (2.4.8)$$

For SIT, the transmitted signal from transmitting antenna index t is given as

$$\mathbf{x}_{SIT,t} = \sqrt{\frac{\beta}{N_{Tx}}}\mathbf{M}_{AddCP}(N_d,L_c-1)[\sqrt{\gamma}\,(\mathbf{1}_{N_c\times 1}\otimes\mathbf{p}_{SIT,t})+\sqrt{1-\gamma}\mathbf{d}_{SIT,t}], \quad (2.4.9)$$

where $\mathbf{p}_{SIT,t} = \mathbf{p}(IDX_{pilot,t})$. Then, in receiving antenna r, the received signal is given as

$$\mathbf{y}_{SIT,r,CP} = \sqrt{\frac{\beta}{N_{Tx}}}\mathbf{X}_{SIT,CP}\mathbf{h}_{r,stacked}+\mathbf{n}_{CP} \qquad (2.4.10)$$

Then, after removing the cyclic prefix, evaluating the cyclic mean, and normalizing with the pilot power γ, CP power allocation factor β, and noting per-Tx-antenna power normalization $1/N_{Tx}$, we obtain

$$\mathbf{y}_{M,SIT,r} = \mathbf{P}_{SIT}\mathbf{h}_{r,stacked}+\sqrt{\frac{N_{Tx}}{\gamma}}\left(\sqrt{1-\gamma}\,\mathbf{D}_M\mathbf{h}_{r,stacked}+\sqrt{\beta^{-1}}\mathbf{n}_{M,r}\right). \qquad (2.4.11)$$

The LS channel estimate is now given as

$$\begin{aligned}\hat{\mathbf{h}}_{LS,r} &= (\mathbf{P}_{SIT}^H\mathbf{P}_{SIT})^{-1}\mathbf{P}_{SIT}^H\mathbf{y}_{M,SIT,r}\\ &= \mathbf{h}_{r,stacked}+(\mathbf{P}_{SIT}^H\mathbf{P}_{SIT})^{-1}\mathbf{P}_{SIT}^H\sqrt{\frac{N_{Tx}}{\gamma}}(\sqrt{1-\gamma}\,\mathbf{D}_M\mathbf{h}_{r,stacked}+\sqrt{\beta^{-1}}\mathbf{n}_{M,r}).\end{aligned}$$
$$(2.4.12)$$

As in the SIT based LS channel estimator for SISO channel (2.2.9), the channel estimation MSE depends on the channel distribution, which significantly complicates the MSE evaluation. Exact analytical solution for certain channel is possible, but here we instead provide the MSE for a unity channel $\mathbf{h}_{i,j,unity} = [1\;\mathbf{0}_{1\times L_c-1}]^T$, where $i = 1,2,...,N_{Rx}$ and $j = 1,2,...,N_{Tx}$. This is to allow us to compare TDMT, SIT and

DDST methods with relatively simple channel estimation MSE equations. The MSE for LMMSE channel estimator with SIT for ideal channel is given as

$$\text{MSE}(\hat{\mathbf{h}}_{LS,SIT,r}) = \frac{N_{Tx}}{\gamma N_c} \left[(1-\gamma)\sigma_d^2 + \frac{\sigma_n^2}{\beta} \right] \text{trace}\{(\mathbf{P}_{SIT}^H \mathbf{P}_{SIT})^{-1}\}. \qquad (2.4.13)$$

For DDST, the transmitted signal from transmitting antenna index t is given as

$$\begin{aligned}
\mathbf{x}_{DDST,i} \;=&\; \sqrt{\frac{\beta}{N_{Tx}}} \mathbf{M}_{AddCP}(N_d, L_c-1) \left[\sqrt{\gamma} \left(\mathbf{1}_{N_c \times 1} \otimes \mathbf{p}_{DDST,t} \right) \right. \\
&\; \left. + \sqrt{1-\gamma} \sqrt{\frac{N_c}{N_c-1}} (\mathbf{I}_N - \mathbf{J}(N_c, N_c)) \mathbf{d}_{DDST,t} \right],
\end{aligned} \qquad (2.4.14)$$

where $\mathbf{p}_{DDST,t} = \mathbf{p}(IDX_{pilot,t})$. Then, in receiving antenna r, the received signal is given as

$$\mathbf{y}_{DDST,r,CP} = \sqrt{\frac{\beta}{N_{Tx}}} \mathbf{X}_{DDST,CP} \mathbf{h}_{r,stacked} + \mathbf{n}_{CP} \qquad (2.4.15)$$

Then, after removing the cyclic prefix, evaluating the cyclic mean of the remaining signal, and normalizing with the pilot power γ, CP power allocation factor β, and per-Tx-antenna power normalization $1/N_{Tx}$, we obtain

$$\mathbf{y}_{M,DDST,r} = \mathbf{P}_{DDST} \mathbf{h}_{r,stacked} + \sqrt{\frac{N_{Tx}}{\beta\gamma}} \mathbf{n}_{M,r}. \qquad (2.4.16)$$

The LS channel estimate is now given as

$$\begin{aligned}
\hat{\mathbf{h}}_{LS,r} \;=&\; (\mathbf{P}_{DDST}^H \mathbf{P}_{DDST})^{-1} \mathbf{P}_{DDST}^H \mathbf{y}_{M,DDST,r} \\
=&\; \mathbf{h}_{r,stacked} + (\mathbf{P}_{DDST}^H \mathbf{P}_{DDST})^{-1} \mathbf{P}_{DDST}^H \sqrt{\frac{N_{Tx}}{\beta\gamma}} \mathbf{n}_{M,r}
\end{aligned} \qquad (2.4.17)$$

and the corresponding MSE over stacked channel vector is

$$\text{MSE}(\hat{\mathbf{h}}_{LS,DDST,r}) = \frac{N_{Tx}}{\beta\gamma N_c} \sigma_n^2 \text{trace}\{(\mathbf{P}_{DDST}^H \mathbf{P}_{DDST})^{-1}\}. \qquad (2.4.18)$$

The extension to LS-LMMSE estimator for TDMT, SIT and DDST cases, is based on the unified LS-LMMSE model (2.3.6) and the MSE is obtained from (2.3.7).

Table 4. *Parameter definitions for unified presentation for the LS-LMMSE channel estimator in the MIMO case. Here, $\sigma_{w,x}^2$ represents the training dependent interference variance.*

x	TDMT	SIT	DDST
$\sigma_{w,x}^2$	$N_{Tx}\sigma_n^2/\lambda$	$N_{Tx}[(1-\gamma)\sigma_d^2+\sigma_n^2/\beta]/(\gamma N_c)$	$N_{Tx}\sigma_n^2/(\beta\gamma N_c)$
z	$\mathbf{y}_{TDMT,p,r}$	$\mathbf{y}_{M,SIT,r}$	$\mathbf{y}_{M,DDST,r}$
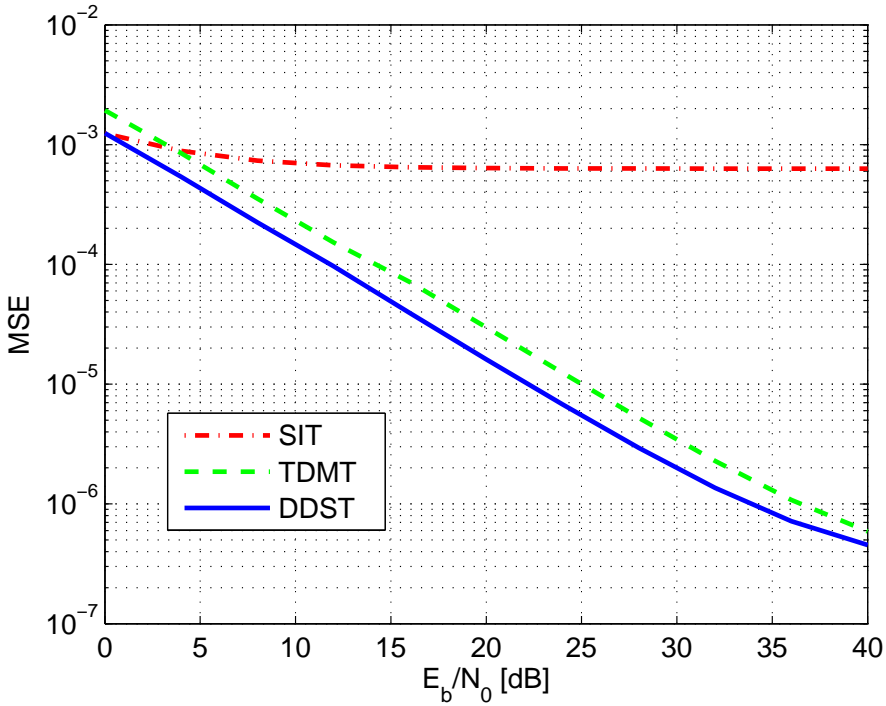
Fig. 2. *Example of the LMMSE channel estimator MSE performance for SIT, DDST and TDMT in a SISO extended ITU Vehicular-A channel. The used parameter values are $\lambda = 0.2$, $N_c = 40$, $N_p = 59$, and $N_{cp} = 57$.*

The values used for in (2.3.6) and (2.3.7) are given in Table 4. Again, we use the power response of the obtained LS channel estimate as the main diagonal for the diagonal channel correlation matrix estimate, $\hat{\mathbf{C}}_{\hat{h},x} = \text{diag}([|\hat{h}_{LS,x,r}(0)|^2|\hat{h}_{LS,x,r}(1)|^2 \cdots |\hat{h}_{LS,x,r}(N_p-1|^2])$ in the channel estimation process, but the analytical MSE is based on the true correlation matrix. Note, that here the effective channel is the stacked version of all channels between transmitting antennae to receiving antenna index r

and possible zero extension, depending on the length of the used pilot vector.

In Fig. 2, we have plotted an example of the MSE performance for SIT, DDST and TDMT using LS-LMMSE channel estimation with pilot power allocation factor $\lambda = 0.2$ in a extended ITU Vehicular-A channel [98]. For this example, we have used parameters $N_c = 40$, $N_p = 59$, $N_{cp} = 57$, and $N_d = 2360$. This leads to $N_{tot} = 2476$ with TDMT and $N_{tot} = 2417$ with DDST and SIT. The N_{cp} is chosen based on the maximum channel delay of the ITU Vehicular-B channel, which is longer than the Vehicular-A channel. The maximum delay of the Vehicular-B channel is 114 samples or 57 symbols in the case of two times oversampled transmission and the maximum delay of the Vehicular-A channel is 78 samples. The N_p value is the smallest prime number larger than the largest expected channel delay in symbols, in other words the smallest prime number with value greater than N_{cp}. Firstly we notice the poor performance of the SIT channel estimator due to the strong self interference. Secondly, we notice that the MSE of the DDST is better than for TDMT given that the same λ is used, as shown in (2.2.15).

3. CHANNEL EQUALIZATION

After channel estimation, the receiver has to utilize the obtained information for channel equalization. In channel equalization, based on the channel estimate, the receiver tries to minimize the distortion of the received symbol or symbol sequence. In some cases the difference between channel equalizer and symbol detector is somewhat blurry because the equalizer may provide in the output hard or soft symbol estimates and sometimes the detection procedure (mapping symbols to hard or soft bits) is an inbuilt function of the equalizer. We talk about equalizing when the received sequence is processed to minimize the effect of the dispersive channel. The equalizer then provides soft symbol estimates that act as an input for the symbol detector, which provides soft bit estimates used by the channel decoder. It is well known that channel decoders perform better with soft bit values [12] and iterative codes, as turbo codes [16] and low-density parity-check codes [43, 70], require soft bit estimates for optimal performance.

For channel equalization the most commonly implemented solutions are the LS and LMMSE approaches [12], as presented for channel estimation. Also maximum likelihood sequence estimators (MLSEs) are popular solutions [41, 112] providing the most likely sequence of symbols in the equalizer output. To simplify the equalizing filters, the equalization can be devided in to feedforward and feedback equalizer branches, where the feedback filter works with detected symbols to alleviate the intersymbol interference after the feedforward filter, as discussed, e.g., in [5, 13, 29, 30]. Controlling these filters adaptively is common choise for wireless communications and includes a vast variety of research work. Interested reader may start with [90, 108] to explore this specific field of study. More recent solutions on the channel equalization rely on iterative signal processing, as discussed, i.e., in [63, 64, 103, 104] and references therein.

The main drawback of SC transmission is the time domain channel equalization complexity. Because the time domain channel equalization complexity increases exponentially with respect to the channel delay, the SC transmission scheme seemed to be doomed since the general acceptance of OFDM transmission and its simple frequency

domain equalization. One solution to decrease the equalization complexity in a SC receiver is to modify the transmitted signal and receiver to also support frequency domain equalization (FDE). Introducing CP for transmitted SC packets and adding discrete Fourier transform (DFT) and inverse discrete Fourier transform (IDFT) signal processing to the receiver, we are capable of equalizing the received signal in frequency domain with only linear increase in the equalization complexity with respect to the channel length. For example, in [14, 31, 42] and references therein the frequency domain equalization combined with SC broadband wireless communications is discussed and the benefits and performance when compared to OFDM are discussed. For more discussion on broadband SC communications, see Appendix D.

It should be emphasized that our studies did not focus on channel equalization algorithms. Instead, we decided to use filter bank (FB) based channel equalizer [52, 123] because we had an efficient implementation ready. This section provides an short introduction to this topic and readers interested can look for [111] and references there in for FB design, and [55, 99, 122] and references there in for equalizer design. The basic principles of FBs are overviewed in Appendix B and further details on the used FB based frequency domain channel estimator are given in Appendix C. Unlike most of the recent research, in our studies we have concentrated on SC transmission with FB based sub-channel wise equalization (SCE).

3.1 Frequency Domain Channel Equalization

For defining the basic equalizer structures, it is better for us to look at frequency domain presentation of the received signal. In general, the frequency domain presentation is obtained per Rx antenna stream by taking the DFT after the removal of CP. We have assumed perfect time and frequency domain synchronization throughout this thesis. Also, the equalizer design is based on 2 times oversampled received signal including the receiver RRC filtering.

For example, the frequency domain presentation for the data signal of TDMT at between transmitting antenna of index t and receiving antenna of index r, is defined as

$$
\begin{aligned}
\mathbf{Y}_{TDMT,t,r} &= \mathrm{DFT}[\mathbf{y}_{d,t,r}] \\
&= \odot(\mathbf{H}_{t,r}, \mathbf{D}_{TDMT,t,r}) + \mathbf{N}_{t,r},
\end{aligned}
\tag{3.1.1}
$$

where $\odot(\mathbf{a}, \mathbf{b})$ defines an element-wise product between vectors $\mathbf{a}$ and $\mathbf{b}$, $\mathbf{H}_{t,r}$ is a

vector containing complex channel weights for each frequency bin and $\mathbf{D}_{TDMT,t,r}$ is the frequency domain presentation of the data vector $\mathbf{d}_{TDMT,t,r}$. The used DFT length is equal or larger than the expected length of the convolution between maximum length channel vector and data vector.

Let us denote by $\mathbf{Y}_k = [Y(1)_k\ Y(2)_k\ \cdots Y(N_{Rx})_k]^T$ the received samples per receiving antenna at frequency bin k. Now, the received signal at frequency bin k can be written as

$$\mathbf{Y}_k = \mathbf{H}_k\mathbf{X}_k + \mathbf{N}_k \tag{3.1.2}$$

where the effective channel matrix is defined as $\mathbf{H}_k = H_{k,RRC}\mathbf{H}_{k,channel}H_{k,RRC}$, modeling the transmit pulse shape filtering, the channel and the receiver pulse shape filtering and $\mathbf{H}_{k,RRC}$ is the frequency domain weighting on carrier k implied by the RRC pulse shape filtering. Here RRC stands for root-raised cosine filters which we have generally assumed to be used as the pulse shape filter in the transmitter and receiver. The MIMO channel matrix in which channel responses from all Tx-Rx antenna pairs have been collected has dimensions $\mathbf{H}_k \propto (N_{Rx}, N_{Tx})$, transmitted data vector has dimensions $\mathbf{X}_k \propto (N_{Tx}, 1)$ and the received signal vector has dimensions $\mathbf{Y}_k \propto (N_{Rx}, 1)$, as does the noise vector $\mathbf{N}_k$. From this presentation it is clear to see, that if we know the effective channel matrix $\mathbf{H}_k$, we can use LS or LMMSE solutions to remove or reduce the effect of the channel distortion and obtain improved estimates of the transmitted samples. In the case of SC transmission, after the equalization we have to convert the frequency domain signal back to time domain for symbol detection.

Assuming that we know the channel response in the receiver, the LS solution for equalizing the received signal is given as

$$\begin{aligned} \hat{\mathbf{X}}_{k,LS} &= (\mathbf{H}_k^H\mathbf{H}_k)^{-1}\mathbf{H}_k^H\mathbf{Y}_k \\ &= \mathbf{X}_k + (\mathbf{H}_k^H\mathbf{H}_k)^{-1}\mathbf{H}_k^H\mathbf{N}_k, \end{aligned} \tag{3.1.3}$$

and the LMMSE solution is typically written as

$$\begin{aligned} \hat{\mathbf{X}}_{k,LMMSE} &= (\mathbf{C}_{X_kX_k}^{-1} + \mathbf{H}_k^H\mathbf{C}_{N_k}^{-1}\mathbf{H}_k)^{-1}\mathbf{H}^H\mathbf{C}_N^{-1}\mathbf{Y}_k \\ &= (\frac{\sigma_{N_k}^2}{\sigma_{X_k}^2}\mathbf{1}_{N_{Tx}} + \mathbf{H_k}^H\mathbf{H}_k)^{-1}\mathbf{H}^H\mathbf{Y}_k, \end{aligned} \tag{3.1.4}$$

where the second line follows from the assumptions of i.i.d data signal, $\mathbf{X}_k$, and noise realization, $\mathbf{N}_k$.

For channel equalizer, the MSE is obtained by taking expectation over the error signal, but now the expectation is taken also over the channel distribution, $\text{MSE}(\hat{\mathbf{X}}_k) = E_{H_k, Y_k, X_k}\left[(\mathbf{X}_k - \hat{\mathbf{X}}_k)^2\right]$. Typically, this leads to analytically challenging derivations for accurate analytical models for practical channels.

Furthermore, in all real life receivers, we do not have the perfect channel knowledge but we have a channel estimate, $\hat{\mathbf{H}}_k$, that is used as it was the ideal channel estimate. This leads to degraded performance in the equalization performance. More discussion on this topic can be found e.g. in [66], where the mismatched LS and LMMSE equalizers based on matrix perturbation theory are studied and in [61], where optimal pilot and data power allocation factors are defined for TDMT and DDST by working in the asymptotic regime (by relaxing the assumption of finite packet sizes).

3.2 Filter Bank Based Channel Equalization

Filter bank (FB) based FDE is an alternative way to perform efficient channel equalization compared to more traditional DFT based methods. We have used the FB based receiver structure because it provides close to ideal linear equalizer performance and has good spectral containment properties (adjacent channel suppression is clearly better than with DFT based solutions). In addition, FB based receiver does not require a cyclic prefix, which can be used to improve the spectral efficiency. As a drawback of FB based equalization one has to consider the higher implementation complexity.

Earlier studies for SC transmission with FB based frequency domain equalizer can be found from [119–121, 123] and references therein. More information for FB based equalizer design for MIMO communications is available from [52] and for SIMO communications in [56].

FB based channel equalizer for SC transmission is built from analysis FB, sub-channel wise equalizer (SCE), and synthesis FB. The analysis FB converts the time domain signal to the frequency domain (similar to the well known DFT operation) and the synthesis FB converts the frequency domain presentation back to time domain (similar to the IDFT operation). The composite sub-channels are recombined in the synthesis FB, which also efficiently realizes the sampling rate reduction by two. The sub-channels in the FB could have wider bandwidth than the expected subcarrier separation in corresponding OFDM design. In any case, since CP is not used in FB systems, the sub-channels are mildly frequency selective and may require more complex equalizer structures when compared to the single tap solution used with OFDM.
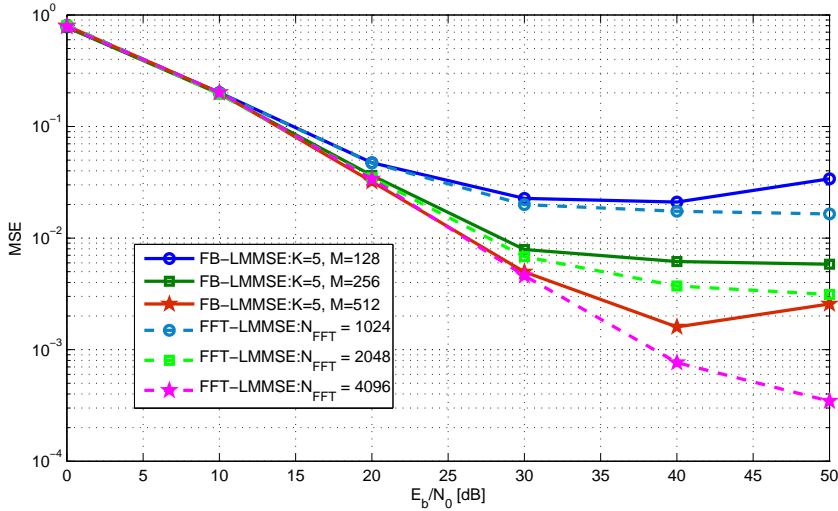
Fig. 3. *Comparison between FB-LMMSE and FFT-LMMSE channel equalizers with different number of sub-channels or FFT bins. The used constellation is 16-QAM, $\gamma = 0.2$, and the channel is ITU Vehicular-A channel. The results are presented only with TDMT to emphasize the differences between FB-LMMSE and FFT-LMMSE with different values of M or N_{FFT}. Using DDST would shift the lines depending on the parametrization but the relative performance between FFT and FB based solutions would remain the same.*

If the OFDM system is properly dimensioned, the effect of the channel is seen as a cyclic convolution between the transmitted signal and the channel response and the desired signal can be perfectly reconstructed by a single tap equalizer in a noiseless channel.

The SCE is performed in the frequency domain with 3-tap complex finite impulse response (FIR) filter for each sub-channel. The equalizers for each Tx branch are designed based on the LMMSE method, presented in [52]. In our simulations we have defined the channel estimates in the time domain and then converted them with DFT operation to the frequency domain. In this case, the DFT operation acts also as an ideal interpolator and extends the shorter time domain channel estimate over the whole frequency range. The channel estimates could also be obtained directly in the frequency domain and after suitable interpolation they could be used for defining the SCE equalizer tap values for each sub-channel.

In Fig. 3 the performance of DFT based and FB based equalizers with different DFT lengths or different number of sub-channels is compared in terms of the channel

equalizer mean squared post-processing error. We notice how in this example the FB based equalizer requires approximately 8 times less sub-channels in the synthesis bank when compared to the number of DFT bins. The frequency domain channel estimate used for FB equalizer has $N_{samples} = 4M$, where M is the number of sub-channels in the synthesis bank. The analysis bank has two times more sub-channels than the synthesis bank because of the oversampling and each sub-channel requires three samples for the 3-tap complex FIR filter. More details on the used FB based SCE can be found from Appendix C.

4. DATA SYMBOL DETECTION

After the channel equalizer the next task in the receiver is to detect the transmitted symbols and bits. Traditionally the receiver would first do hard detection of the transmitted symbols and then map these symbol estimates into a hard bit stream. Then the channel decoder would try to decode the transmitted data word based on the received hard code word.

Nowadays, a more common approach is to evaluate the soft bit estimates from the equalizer output. The equalizer output itself is a soft symbol estimate of the transmitted symbol vector. Before the soft symbols to bits mapping we have to properly scale the received signal power depending on the used training method and remove the pilot signal. Especially in the case of SIT we have to carefully remove the pilot signal in order to minimize the possible interference on top of the data symbols.

4.1 Power Scaling for Different Training Methods

Let us next briefly define the received, power scaled soft symbol estimates for different training methods. For TDMT, the soft data symbol estimates after power scaling are defined as

$$
\begin{aligned}
\hat{\mathbf{d}}_{TDMT} &= \tfrac{1}{\sqrt{1-\gamma}}\hat{\mathbf{x}}_{TDMT} \\
&= \mathbf{E}\mathbf{d} + \tfrac{1}{\sqrt{1-\gamma}}\mathbf{W}\mathbf{n}_d,
\end{aligned}
\tag{4.1.1}
$$

where $\mathbf{E}$ is a matrix includes the residual error of the channel estimation error after channel equalization and $\mathbf{W}$ is the channel equalizer weight matrix, which is based on distorted channel estimates $\hat{\mathbf{H}}$ obtained from the channel estimator. The residual error matrix is defined as $\mathbf{E} = \mathbf{W}\mathbf{H} = \mathbf{W}(\mathbf{H} + \Delta - \Delta) = \mathbf{W}\hat{\mathbf{H}} - \mathbf{W}\Delta$, where Δ is a matrix modeling the channel estimation error of the used channel estimator. In ideal case, $\mathbf{E} = \mathbf{I}$, and no residual interference on top of the data sequence is present in the

equalizer output. The channel estimation error matrix is often assumed to be Gaussian or complex Gaussian diagonal matrix, but it can also include estimation error components from non-Gaussian distributions, common scaling component caused by PA distortion, per channel index scaling components, or correlation components between error values.

For SIT, the definition of soft data symbols is not as straightforward as with TDMT or DDST. With SIT, we encounter a problem when we want to remove the pilot sequence added on top of the data sequence. If the channel estimate is perfect ($\mathbf{E} = \mathbf{I}$) and we use LS equalizer, we can completely remove the pilot sequence, but otherwise there is a additional interference term on top of the data signal caused by non-ideal removal of the pilot signal. Let us approximate the output of the channel equalizer for SIT as

$$
\begin{aligned}
\hat{\mathbf{x}}_{SIT} &= \mathbf{E}\mathbf{x}_{SIT} + \mathbf{W}\mathbf{n} \\
&= \mathbf{E}\sqrt{\beta}\left[\sqrt{\gamma}(\mathbf{I}_{N_c \times 1} \otimes \mathbf{p}_{SIT}) + \sqrt{1-\gamma}\mathbf{d}\right] + \mathbf{W}\mathbf{n},
\end{aligned}
\tag{4.1.2}
$$

Now, because we want to detect the data symbols, we have to remove the pilot signal from the received symbols. At this point, our best guess is to remove correctly scaled cyclic pilot from the received signal, leaving us with

$$
\begin{aligned}
\hat{\mathbf{d}}_{SIT} &= \frac{1}{\sqrt{\beta(1-\gamma)}}\left[\hat{\mathbf{x}}_{SIT} - \sqrt{\beta\gamma}(\mathbf{I}_{N_c \times 1} \otimes \mathbf{p}_{SIT})\right] \\
&= (\mathbf{E}-\mathbf{I})\sqrt{\frac{\gamma}{1-\gamma}}(\mathbf{I}_{N_c \times 1} \otimes \mathbf{p}_{SIT}) + \mathbf{E}\mathbf{d} + \frac{1}{\sqrt{\beta(1-\gamma)}}\mathbf{W}\mathbf{n} \\
&= \Delta\mathbf{p}_{SIT} + \mathbf{E}\mathbf{d} + \frac{1}{\sqrt{\beta(1-\gamma)}}\mathbf{W}\mathbf{n}.
\end{aligned}
\tag{4.1.3}
$$

From this kind of approximation, it is clear that in addition to distortion of the data symbols, the relatively poor channel estimate obtained with SIT also degrades the pilot removal accuracy, causing a significant pilot error term $\Delta\mathbf{p}_{SIT}$.

With DDST, we do not have the same problem as with SIT, due to the orthogonality between $\mathbf{d}_{DDST}$ and $\mathbf{p}_{DDST}$. In other words, the pilot signal is completely removed by removing the cyclic mean of the received signal. Thus, that the data symbol estimate with DDST is defined as

$$
\begin{aligned}
\hat{\mathbf{d}}_{DDST} &= \sqrt{\frac{N_c-1}{\beta(1-\gamma)N_c}}\left[\mathbf{I}_{N_d} - \mathbf{J}(N_c,N_c,L_c)\right]\hat{\mathbf{x}}_{DDST} \\
&= \left[\mathbf{I}_{N_d} - \mathbf{J}(N_c,N_c,L_c)\right]\left(\mathbf{E}\mathbf{d}_{DDST} + \sqrt{\frac{N_c-1}{\beta(1-\gamma)N_c}}\mathbf{W}\mathbf{n}\right).
\end{aligned}
\tag{4.1.4}
$$

As discussed in Chapter 2, the problem with symbol detection when considering DDST is the time varying nature of the self interference term $\mathbf{P}_d = \mathbf{J}(N_c, N_c, L_c)\mathbf{d}_{DDST}$, which can be considered as time varying modulation per packet basis. The simplest solution to this problem is an iterative symbol detector which based on the symbol estimates generates a cyclic mean estimate that is used to obtain new symbol esti- mates. The cyclic mean estimate can be defined before or after the channel decoding, or both.

The optimal maximum likelihood detector becomes far too complex for symbol de- tection with DDST as the number of cycles or the number of symbols that participate on the cyclic mean evaluation increases. The detector should detect N_p symbols simultaneously to optimally incorporate the distorting mean component in the de- tection. Fortunately, because the data signal has zero mean, the distribution of the error component has its maximum on the zero value (as we will see in Chapter 5). Therefore, using the apriori assumption of zero interference on top of the desired data symbols leads to relatively good initial estimate of the transmitted data symbols. One additional unpublished solution, defined by M.Sc. Jukka Talvitie, is to evaluate the LS solution incorporating the mean term over each set of N_p symbols in the sym- bol detection. This solution provides very good initial symbol estimates, if iterative reception algorithms are not desired.

In Fig. 4, a block diagram of the MIMO transmitter and receiver is shown, including a possible feedback loop for iterative data symbol/bit detection with DDST and for SIT receiver. With SIT there is a feedback loop from symbol detector or/and from soft channel decoder to the channel estimator, because the channel estimation perfor- mance of SIT scheme can be significantly improved by using knowledge of the data symbols to reduce the interference on top of the training signal or using combined information of training and data signal for channel estimation. For example, in the case of LS channel estimate for SIT given in (2.2.8), the new formulation for training matrix when using both pilot and data estimates for channel estimation would be

$$\hat{\mathbf{X}}_{M,SIT} = \mathbf{P}_{SIT} + \frac{\sqrt{1-\gamma}}{\sqrt{\gamma}}\hat{\mathbf{D}}_M, \qquad (4.1.5)$$

where $\hat{\mathbf{D}}_M$ is the cyclic mean of the latest data symbols based on either soft or hard estimates, and the new channel estimate would be defined as

$$\hat{\mathbf{h}}_{LS,SIT,iterative} = (\hat{\mathbf{X}}_{M,SIT}^H \hat{\mathbf{X}}_{M,SIT})^{-1}\hat{\mathbf{X}}_{M,SIT}^H \mathbf{y}_{M,SIT}. \qquad (4.1.6)$$
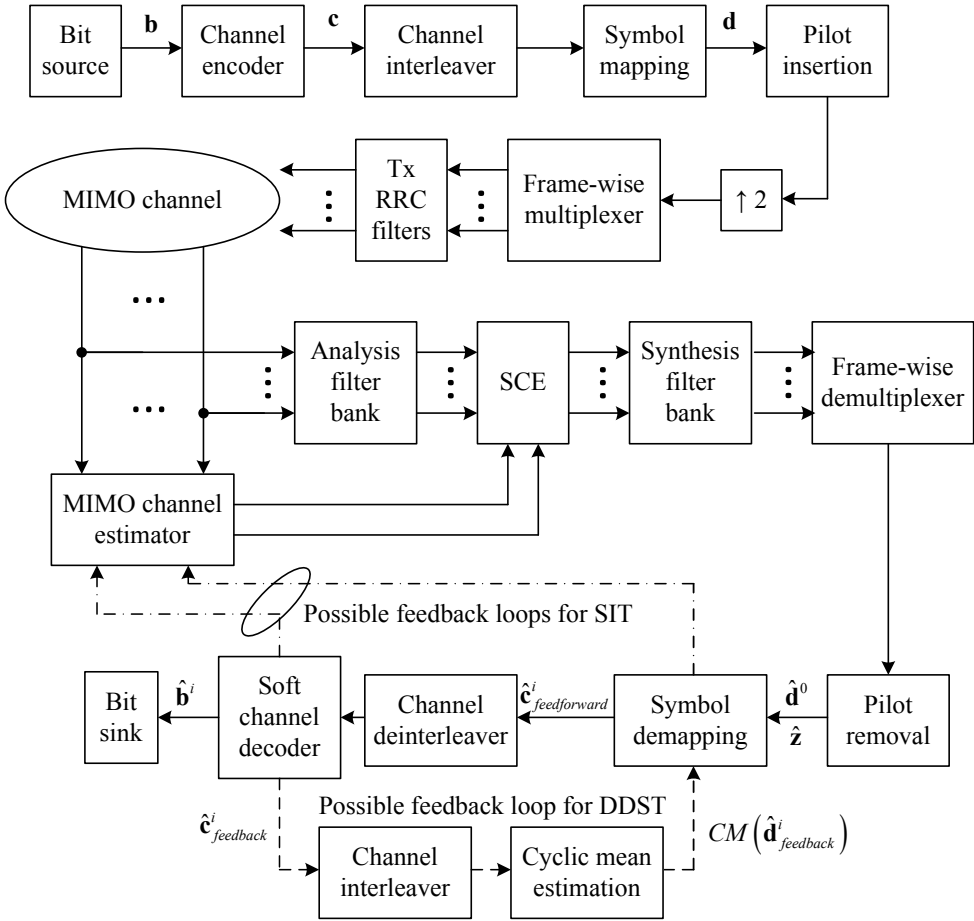
Fig. 4. *Transmitter and receiver model for MIMO transmission. The receiver pulse shape filtering is included into the LMMSE based sub-channel wise equalizer (SCE) and the synthesis filter bank efficiently incorporates the two times down sampling of the received signal.*

We will focus on the iterative symbol detection with DDST. Discussion on iterative receiver structures for SIT can be found, for example, from [37, 100]. We have not considered including the channel estimation process in the iterative loop with DDST because there is no interference from the data symbols on the known pilot symbols.

4.2   Iterative Cyclic Mean And Data Symbol Estimation

In this section, we provide a simplified excursion to iterative data symbol detection with DDST. A more detailed description, including a symbol level limiting operation, is described in [P4]. The derivations and discussion of Chapter 2 are related to the block `MIMO channel estimator` presented in Fig. 4. Chapter 3 discussed in general the frequency domain equalization related to the procedures performed in the block `SCE`, corresponding to sub-channel equalizer design used in filter bank based equalizer.

A simple means of improving the symbol detection with DDST based on the hard symbol estimates before channel decoding was proposed on [46]. There it was shown that already one iteration using uncoded hard symbol estimates significantly improves the performance. Based on our practical experience [P4], also in the case of using feedback from the channel decoder, it is better to first estimate the missing cyclic mean component based on received symbols before generating soft coded bit estimates for the channel decoder.

In Fig. 4, after the frame-wise multiplexer, we have the `Pilot removal` block. Inside this block, the pilot signal is removed, the data signal is power scaled and we obtain the data symbol estimate $\hat{\mathbf{d}}_x$, defined in either (4.1.1), 4.1.3), or (4.1.4), corresponding to training scheme TDMT, SIT, or DDST, respectively.

The signal power level might be affected by, e.g., predistortion, saturation in the power amplifier (PA), or LMMSE equalization in low SNR scenarios. In the case of SIT, the received sequence power needs to be normalized to the correct level, before removing the known training signal. TDMT and DDST do not require power normalization for training signal removal, but for correct symbol detection all schemes need to be on the desired average power level. For DDST, the pilot signal is removed by removing the cyclic mean of the received signal. Note that this operation nulls certain frequency bins and therefore also removes the noise and interference located in these bins from the received signal.

Here $\hat{\mathbf{z}}$ represents our estimates of the DDST data symbol sequence after receiver side frame-wise demultiplexer. We generate hard symbol estimates based on $\hat{\mathbf{z}}$, calculate their cyclic mean and add it to $\hat{\mathbf{z}}$, to obtain initial symbol estimates $\hat{\mathbf{d}}^0$, given as

$$\hat{\mathbf{d}}^0 = \hat{\mathbf{z}} + CM(HSM(\hat{\mathbf{z}})), \tag{4.2.1}$$

where $HSM(\cdot)$ stands for hard symbol mapper and $CM(\mathbf{x}) = \mathbf{J}(N_c, N_c, L_c)\mathbf{x}$ provides

the cyclic mean of $\mathbf{x}$. Here superscript 0 points out that these symbol estimates are obtained before coded feedback. This idea was presented in [46], and we use it before the first run through the decoder.

We start the iterative reception process by using $\hat{\mathbf{d}}^0$ to generate soft coded bit estimates $\hat{\mathbf{c}}^0_{feedforward}$. These are then provided to the soft-input soft-output (SfISfO) decoder from which we obtain updated soft coded bit estimates, $\hat{\mathbf{c}}^i_{feedback}$, and soft data bits, $\hat{\mathbf{b}}^i$, for bit error evaluation. In our notation, depicted also in Fig. 4, after the channel decoder we start a new iteration. The updated coded bits, $\hat{\mathbf{c}}^i_{feedback}$, from the output of the channel decoder are fed back for soft cyclic mean evaluation that is used to improve the soft symbol estimates. In iteration i, the improved soft symbol estimates are defined as

$$\hat{\mathbf{d}}^i = \hat{\mathbf{z}} + CM(\hat{\mathbf{d}}^i_{feedback}), \qquad (4.2.2)$$

where $\hat{\mathbf{d}}^i_{feedback}$ are the soft symbols defined based on latest soft coded bits $\hat{\mathbf{c}}^i_{feedback}$. From the improved soft symbol estimates, $\hat{\mathbf{d}}^i$, we generate new coded bit estimates, $\hat{\mathbf{c}}^i_{feedforward}$, for the channel decoder to generate new soft data bit estimates and possibly continue detection iterations.

The soft symbol estimates are based on the latest soft bit estimates $\hat{\mathbf{c}}^i_{feedback}$, which are equal to the log-likelihood presentation of the a posteriori probabilities obtained from the soft decoder. The soft symbols are given by equation

$$\hat{d}^i_{\nu,feedback} = \sum_{i=1}^{|\Omega_D|} d_i p(d_i | \hat{\mathbf{c}}^i_{\nu,feedback}), \ 0 \le \nu \le N-1, \qquad (4.2.3)$$

where $|\Omega_D|$ gives the number of symbols in alphabet Ω_D, ν is a symbol index, $\hat{\mathbf{c}}^i_{\nu,feedback}$ are the soft bit estimates related to the νth symbol, and $p(d_i | \hat{\mathbf{c}}^i_{\nu,feedback})$ is the probability of a symbol d_i, given the latest soft coded bit estimates $\hat{\mathbf{c}}^i_{\nu,feedback}$. The probability of a symbol d_i is defined as

$$p(d_i | \hat{\mathbf{c}}^i_{\nu,feedback}) = 2^{-q} \prod_{j=1}^{q} \left[1 + \bar{c}_{d_i}(j) \tanh\left(\frac{\hat{c}^i_{\nu,feedback}(j)}{2} \right) \right], \qquad (4.2.4)$$

where q is the number of bits per symbol, $\bar{c}_{d_i}(j) \in [-1, +1]$ is the jth bit of the hypothesis d_i, and $\hat{c}^i_{\nu,feedback}(j)$ is the log-likelihood presentation of the a posteriori probability related to the jth bit of the νth symbol in the ith iteration, given as
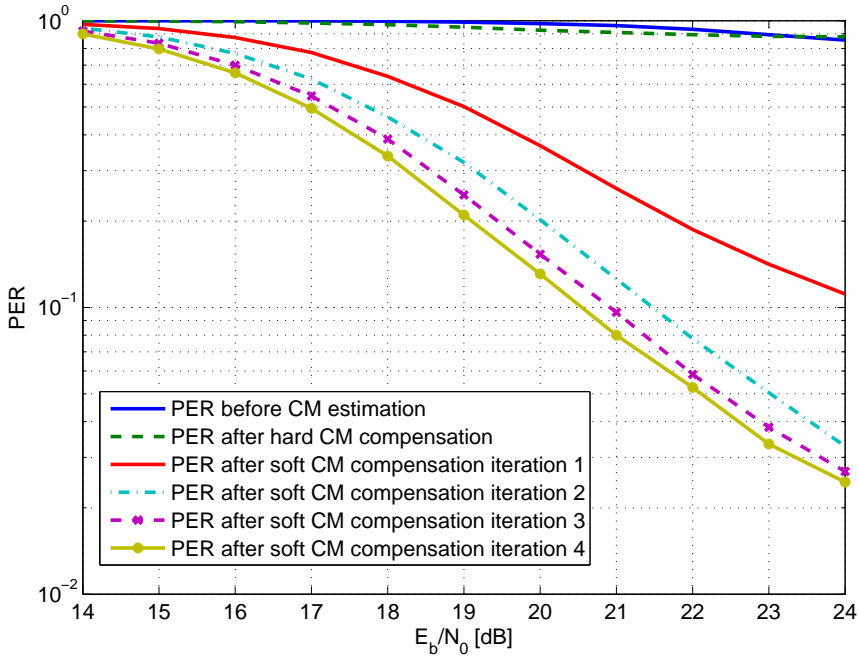
Fig. 5. *PER performance for 64-QAM with DDST, LMMSE channel estimation and equalization, coding rate $R = 3/4$, $\gamma = 0.1$, $N_c = 40$, $N_p = 59$, and four soft detection iterations.*

$$\hat{c}^i_{v,feedback}(j) = \log \left(\frac{P_{app}(c^i_{v,feedback}(j) = 1)}{P_{app}(c^i_{v,feedback}(j) = 0)} \right). \qquad (4.2.5)$$

We have also normalized the variance of the soft symbol vector, $\hat{\mathbf{d}}^i$, to be equal to unity. This improves the feedback performance when the soft bit estimates have very low reliability.

Assuming that the SISO decoder has been able to reduce the number of bit errors in the detected bit sequence, the estimate of the self interference term, $\hat{\mathbf{p}}^i_d = -CM(\hat{\mathbf{d}}^i)$, at each increasing iteration is an improved estimate of the cyclic mean. Therefore, the symbol error rate and bit error rate are decreasing functions with respect to the iteration number. This holds if the initial symbol estimates are good enough to initiate the convergence of the receiver.

Based on our results, it is better not to use the extrinsic information obtained from the channel decoder as apriori information in the soft symbols-to-bits mapping, if this

information is already used to improve the cyclic mean estimate. This is probably because we are using the same information twice inside the same loop, thus losing the independence of the apriori information. We can use it as apriori information if we do not improve the cyclic mean, but based on our studies this does not provide as good iterative gain in the receiver. This could be because of the error averaging nature of the cyclic mean computation. For future studies, testing other bits to symbol mappings instead of Gray mapping would be an interesting addition to this topic.

In Fig. 5, the PER performance of the iterative receiver with four soft symbol detection and CM estimation iterations is shown, including the PER with or without hard symbol based CM estimation. The used modulation is 64-QAM, coding rate $R = 3/4$, and number of cyclic copies is $N_c = 40$. The used channel model is extended ITU Vehicular A channel model [98]. We can see significant improvement in the PER performance with the iterative soft symbol detector. Especially higher constellations, that are more sensitive to the self interference, gain significantly from using the iterative detector. In [P4], the performance of this iterative detector combined with limiter error estimation is evaluated in detail and performance comparisons with TDMT are provided.

5. PERFORMANCE ANALYSIS FOR DDST

In this chapter, we derive the analytic symbol error rate (SER) and bit error rate (BER) for the DDST based transmission in a Gaussian channel, for real pulse amplitude modulation (PAM) constellations. First, in Section 5.1, we derive the multinomial distribution for the cyclic mean component (data dependent pilot). Then, in Section 5.2, we consider a purely Q-function based presentation which basically models the error probability based on the Gaussian cumulative density function (CDF) and the minimum distance between two symbols in a constellation, d. In Section 5.3, we derive the mutual information (MI) between transmitted and received symbols with DDST. Based on these results and by using the Gaussian presentation for soft information obtained from the MI [102], we can obtain equivalent noise variance for soft data bits obtained from DDST data symbols and then use the Q-function for the soft bits to obtain related BER results. Finally, in Section 5.4, we provide numerical comparison between simulated and analytical results.

The Q-function is defined for the standard Gaussian distribution with variance $\sigma^2 = 1$ and expected value $\mu = 0$, defined as

$$Q(t) = P(X > t) = 1 - F(t) = \frac{1}{\sqrt{2\pi}} \int_t^\infty \exp(-x^2/2)dx. \qquad (5.0.6)$$

When the variance or the mean of the Gaussian distribution has other values, the result is obtained by change of variables, defined as

$$Q(\frac{t' - \mu}{\sigma}) = P(X > t'). \qquad (5.0.7)$$

5.1   Multinomial Distribution for the Cyclic Mean Component

For us to analytically model the SER or BER of a DDST transmission, we have to derive a model for the distribution of the cyclic mean (CM) component. We assume

that the symbol sequence length is equal to $N_d = N_c N_p$, where N_c is the number of cyclic copies and N_p is length of one cycle. The basic model for a DDST signal is given as

$$\mathbf{y} = \sqrt{1-\gamma}(\mathbf{d} - \mathbf{z}) + \sqrt{\gamma}\mathbf{p}_c + \mathbf{n}, \tag{5.1.1}$$

where $\mathbf{y}$ is a symbol sequence containing the data sequence $\mathbf{d}$, CM vector $\mathbf{z} = \mathbf{p}_d$ (data-dependent pilot sequence), the known cyclic pilot sequence $\mathbf{p}_c = \mathbf{1}_{N_c \times 1} \otimes \mathbf{p}$, and the Gaussian noise vector $\mathbf{n}$. Here, γ is the pilot power allocation factor. For clarity, in this chapter we have chosen to use notation $\mathbf{z}$ for the cyclic mean component.

The known pilot sequence plays no role in our future derivations, so we drop it out. Of course, the actual channel estimation and related analysis is an interesting and active area of research, but we have chosen to derive the BER on a AWGN channel so the pilot sequence is not required nor do the results depend on any channel estimation or equalization algorithm. Our results, however, can be used to predict the BER with certain pilot sequence $\mathbf{p}$ and allocated power γ after one has analyzed the MSE in the output of the used channel equalizer, based on the channel estimates obtained from the known pilot sequence, simply by replacing the σ_n^2 used here with the properly scaled post-processing MSE of the used channel equalizer. Also, we have derived these models for PAM but the results are easily generalized to quadrature amplitude modulation (QAM), because of the independency between real and imaginary axes.

The CM of a sequence $\mathbf{d}$ is defined as

$$\mathbf{z} = \mathbf{J}(N_c, N_c, N_p)\mathbf{d} = \left(\frac{1}{N_c}\mathbf{1}_{N_c \times N_c} \otimes \mathbf{I}_{N_p}\right)\mathbf{d}, \tag{5.1.2}$$

or equivalently formulated as a summation per sample,

$$z(\kappa + \iota N_p) = 1/N_c \sum_{j=0}^{N_c-1} d(\kappa + jN_p), \tag{5.1.3}$$

for $\iota = [0, 1, \ldots, N_c - 1]$ and $\kappa = [0, 1, \ldots, N_p - 1]$. Thus, in DDST, on top of each data symbol, we have an average component, defined over N_c symbols and each DDST frame carries N_p different CM values. This also implies that each cyclic mean value z depends on N_c symbols and that if we know z and $N_c - 1$ symbols, we can discover the missing symbol (at least in high SNR case).

The probability of a certain combination of N_c symbols follows the multinomial distribution

$$p(n_1, n_2, \ldots, n_k; n, p_1, p_2, \ldots, p_k) = \begin{cases} \dfrac{n!}{n_1!n_2!\ldots n_k!} p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}, & \text{when } \displaystyle\sum_{i=1}^{k} n_i = n \\[2mm] 0 & \text{otherwise,} \end{cases}$$

(5.1.4)

where n_i is the number of observations of a certain i^{th} constellation point (symbol value) on a real (or imaginary) axis, p_i is the probability of that constellation point and in our case $n = N_c$ is the number of observations in total per CM value. Here k is the number of constellation points per real or imaginary axis. In our case, because all symbols are equally probable, $p_i = 1/k$ for all i. One interesting topic for further research would be to incorporate the symbol probabilities from the feedback loop, defined in (4.2.4), to the multinomial distribution to obtain updated CM probability distribution to be used in the soft detector.

To get the actual probability of a certain CM value, one has to add together all the probabilities of different combinations having that specific CM value. Thus, the probability distribution function of the CM distribution is defined as

$$\begin{aligned} p(z) &= \sum_{\mathbf{d} \in \Omega_\mathbf{d}} \delta\left(z - \frac{1}{N_c} \sum_{i=1}^{N_c} \mathbf{d}(i)\right) p(n_1, n_2, \ldots, n_k; n, p_1, p_2, \ldots, p_k), \\ &= \sum_{\mathbf{d} \in \Omega_\mathbf{d}} \delta\left(z - \frac{1}{N_c} \sum_{j=1}^{k} n_j d_j\right) \frac{n!}{n_1!n_2!\ldots n_k!} \left(\frac{1}{k}\right)^{N_c}, \end{aligned}$$

(5.1.5)

where $d_i, i = [1, 2, \ldots, k]$ represent the possible symbol values per real or imaginary axis, $\mathbf{d} = [d(0), d(1), \ldots, d(N_c - 1)]^T$ is vector of dimensions $(N_c \times 1)$ containing n_1 d_1's, n_2 d_2's, ..., n_k d_k's and $\delta(t) = 1$, if $t = 0$, and $\delta(t) = 0$, otherwise. With high number of cyclic copies, the distribution of the CM value tends toward the Gaussian distribution, as expected based on the central limit theorem. So one can also choose to use Gaussian distribution for modeling the CM distribution, when the number of cyclic copies is high enough. Then the appropriate choice for Gaussian density function approximating the CM distribution $p(z)$ is

$$p(z) \approx p_G(z) = \frac{1}{\sqrt{2\pi\sigma_z^2}} \exp\left(\frac{-z^2}{2\sigma_z^2}\right),$$

(5.1.6)
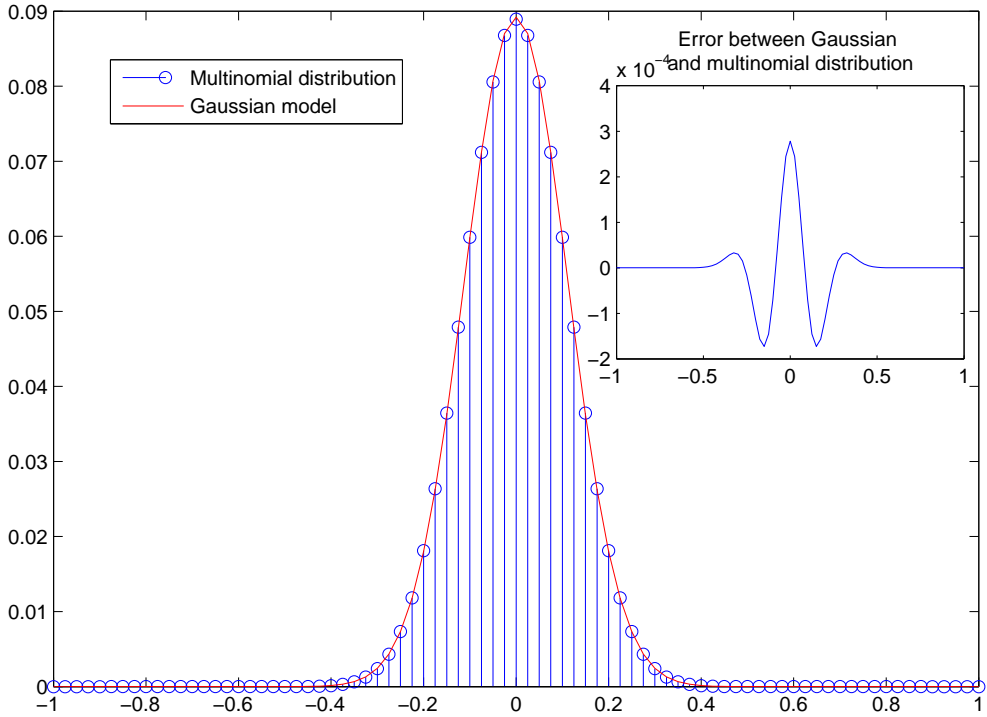
Fig. 6. *Example of the multinomial distribution for 2-PAM constellation and its Gaussian approximation with $N_c = 80$ and $\gamma = 0.1$.*

where the variance equals $\sigma_z^2 = \sigma_d^2/N_c$. These models for the CM distribution were introduced in [P4], but without the exact presentation for the CM probability distribution. In Fig. 6, an example of the multinomial distribution for 2 level pulse-amplitude modulation (2-PAM), is given with the Gaussian approximation. We know the possible cyclic mean values and from (5.1.6) we obtain approximated probability values. The Gaussian approximate is especially important for large constellations or large number of cyclic copies, because the evaluation of the exact distribution is numerically exhaustive operation.

5.2   SER and BER Analysis Based on the Q-function

Let us assume that we are using some PAM modulation with $k = 2^q$ elements with uniform distribution, where each symbol carries q bits. The constellation points are given by a set $\Omega_D = [-2^q + 1, \ -2^q + 1 + \zeta, \ldots, 2^q - 1]$ and each of them has a probability $P(D = d) = 1/2^q$. The distance between symbols is defined as ζ and the

distance from a symbol to a decision threshold as $a = \zeta/2$. The variance of the constellation is defined as $\sigma_d^2 = \sum_{d \in \Omega_D} |d|^2 P(D = d)$. The set of possible CM values related to this constellation is given as $\Omega_Z = [-2^q + 1, \ -2^q + 1 + \zeta/N_c, \ldots, 2^q - 1]$ and the size of the set is equal to $N_Z = |\Omega_Z| = (2^q - 1)N_c + 1$. Different power scalings, e.g. with factor $\sqrt{1 - \gamma}$, affect the values in sets Ω_D and Ω_Z, but the probabilities related to these values are not affected.

Let us remind of the symbol error probabilities with traditional PAM constellation. For an outer symbol, which has only one neighbor in a real constellation, a symbol error occurs with probability $P(\hat{d}_o \neq d) = Q(a/\sigma_n)$, where the noise variance is given by σ_n^2. For an inner symbol, which has two neighbors in a real constellation, a symbol error occurs with probability $P(\hat{d}_i \neq d) = 2P(\hat{d}_o \neq d)$.

Next we define the SER probability for DDST. Because the distribution $p(z)$ is symmetric, we go through only the indexes belonging to the positive error terms and include the negative values in (5.2.1). This is valid, because the negative and positive values with the same absolute value have the same probability. The outer symbol error probability is given as

$$
\begin{aligned}
P(\hat{d}_0 \neq d) \quad &= p(z|z = 0)Q(a/\sigma_n) \\
&+ \sum_{z \in \{0 < \Omega_Z \leq a\}} p(z)\left[Q\left((a+z)/\sigma_n\right) + Q\left((a-z)/\sigma_n\right)\right] \\
&+ \sum_{z \in \{a < \Omega_Z\}} p(z)\left[Q\left((a+z)/\sigma_n\right) + \left(1 - Q\left((z-a)/\sigma_n\right)\right)\right],
\end{aligned}
\tag{5.2.1}
$$

where the probability of the cyclic mean value being equal to zero, $p(z|z = 0)$, is equal to zero when N_z is even (N_c is odd).

The term $(1 - Q((z-a)/\sigma_n))$ appearing when $z > a$ corresponds to the case that the CM component itself causes an symbol error for an outer symbol, and actually a noise component to the right direction, with absolute value larger than $z - a$ but smaller than $z + a$, leads to a correct symbol estimate.

For example, the symbol error probability for 2-PAM corresponds now to $P(\hat{d} \neq d) = P(\hat{d}_o \neq d)$ and for 4-PAM it corresponds to $P(\hat{d} \neq d) = 0.5P(\hat{d}_o \neq d) + 0.5P(\hat{d}_i \neq d) = 1.5P(\hat{d}_o \neq d)$. From these estimates the probability of bit error is estimated simply by dividing the symbol error probability with the number of bits per symbol, if we assume Gray mapping to be used. This approximation is valid only for relatively large number of cyclic copies, $N_c \geq 10$. This is because with small values of N_c, the

probability of large CM values is relatively large and large CM values can cause symbol errors other than with the closest neighbor, thus increasing the number of bit errors.

In our studies, the SER of the DDST based transmission converges to a certain threshold. This is caused by the CM values whose absolute value is larger than a in the used constellation. We call the SER (BER) value in the error floor region as residual SER (BER). A simple approximate of the expected residual symbol error probability can be given based on the probability that the CM component exceeds the decision bounds, given for real constellations as

$$P(\text{residual SER}) = (2k-2)/k \sum_{z \in \Omega_Z} p(z)u(z-a), \qquad (5.2.2)$$

and the bit error probability, under Gray mapping assumption, as

$$P(\text{residual BER}) = P(\text{residual SER})/q, \qquad (5.2.3)$$

where $u(t)$ is the step function defined as $u(t) = 1$, if $t \geq 0$, and $u(t) = 0$, otherwise, $k = 2^q$ is the size of the constellation and q is the number of bits carried per symbol.

5.3 BER Analysis based on the Mutual Information

The mutual information between two variables is a widely used measure for determining the reliability of the information obtained in the soft estimates of the transmitted data. Typically, the mutual information for two variable case in a Gaussian channel is defined as $I(D;Y) = H(Y) - H(Y|D) = H(Y) - H(N)$, where $H(Y)$ is the entropy of the received random variable Y and $H(Y|D) = H(N)$ is the conditional entropy of the received random variable given the transmitted random variable D, which in Gaussian channel corresponds to the entropy of the noise component $H(N)$.

In the case with DDST, if we consider the CM component as noise, then the received signal can be modeled as $y = d - z + n = d + w$. In our derivations we exclude the special cases $N_c < 2$ and $N_d = 1$, which would lead to unwanted result of having $z = d$. Thus, the cyclic mean would be deterministic function of the data symbol and more importantly the transmitted signal would not convey any information. It should be noted that the special case $N_c = N_p = N_d = 1$ was used as a misleading example in [17], leading to the unwanted result with DDST. When $N_c >= 2$, the

cyclic mean component z does depend on the value of d, but is defined by all the other symbols participating in the evaluation of the cyclic mean. Typically, in SIT or DDST parametrization the value $N_c \geq 10$ is used. The probability distribution function of the interference term, given the probability density function of the zero mean Gaussian noise f_n, is defined as

$$
\begin{aligned}
f(w) &= \int_\lambda f_n(\lambda) p_z(w - \lambda) d\lambda \\
&= \int_\lambda \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(\frac{-\lambda^2}{2\sigma_n^2}\right) \sum_{k=1}^{N_Z} \delta((w - \lambda) + z(k)) p(z(k)) d\lambda \\
&= \frac{1}{\sqrt{2\pi\sigma_n^2}} \sum_{k=1}^{N_Z} \exp\left(\frac{-(w + z(k))^2}{2\sigma_n^2}\right) p(z(k)).
\end{aligned}
\tag{5.3.1}
$$

Thus, the mutual information with DDST without apriori knowledge of the CM component is defined as

$$
\begin{aligned}
I(D;Y) &= H(Y) - H(Y|D) = H(Y) - H(W) \\
&= -\int_{y \in \Omega_Y} f(y) \log_2(f(y)) dy + \int_{w \in \Omega_W} f(w) log_2(f(w)) dw.
\end{aligned}
\tag{5.3.2}
$$

In the case of three random variables, one can write the mutual information between the variables as $I(Y;D,Z) = I(D;Y|Z) - I(D;Y) = I(D;Z|Y) - I(D;Z) = I(Y;Z|D) - I(Y;Z)$ [73], [71]. In this paper, we are interested in the conditional mutual information

$$
\begin{aligned}
I(D;Y|Z) &= H(D|Z) + H(Y|Z) - H(Y,D|Z) \\
&= H(Y|Z) - H(Y|D,Z),
\end{aligned}
\tag{5.3.3}
$$

which corresponds to the case that our receiver has apriori knowledge of the cyclic mean component and uses this information in the detection. Assuming now that Y represents the received noisy random signal, D corresponds the uniformly distributed data sequence, and Z represents the cyclic mean component, the conditional entropies

are defined as

$$
\begin{aligned}
H(Y|Z) &= \sum_{z \in \Omega_Z} H(Y|z) p(z) \\
&= - \sum_{z \in \Omega_Z} \int_{y \in \Omega_Y} f(y|z) \log_2[f(y|z)] dy \, p(z) \\
&= - \sum_{z \in \Omega_Z} \int_{y \in \Omega_Y} \sum_{d \in \Omega_D} f(y|d,z) p(d) \log_2 \left[\sum_{d \in \Omega_D} f(y|d,z) p(d) \right] dy \, p(z),
\end{aligned}
$$
(5.3.4)

and

$$
\begin{aligned}
H(Y|D,Z) &= \sum_{d \in \Omega_D} \sum_{z \in \Omega_Z} H(Y|d,z) p(d,z) \\
&= - \sum_{d \in \Omega_D} \sum_{z \in \Omega_Z} \int_{y \in \Omega_Y} f(y|d,z) \log_2[f(y|d,z)] dy \, p(d,z) \\
&= - \sum_{d \in \Omega_D} \sum_{z \in \Omega_Z} \int_{y \in \Omega_Y} f(y|d,z) \log_2[f(y|d,z)] dy \, p(z|d) \, p(d).
\end{aligned}
$$
(5.3.5)

The conditional probability distribution, $p(z|d)$, is defined for the case $d(0) = d$ as

$$
\begin{aligned}
p(z|d) &= p(z|d(0) = d) \\
&= \sum_{\substack{\mathbf{d} \in \Omega_{\mathbf{D}}, \\ d(0) = d}} \delta \left(z - \frac{1}{N_c} \sum_{i=0}^{N_c-1} d(i) \right) p(n_1, n_2, \dots, n_k; n = N_c - 1, p_1, p_2, \dots, p_k) \\
&= \sum_{\substack{\mathbf{d} \in \Omega_{\mathbf{D}}, \\ d(0) = d}} \delta \left(z - \frac{1}{N_c} \left(d + \sum_{j=1}^{k} n_j d_j \right) \right) \frac{(N_c - 1)!}{n_1! n_2! \dots n_k!} \left(\frac{1}{k} \right)^{N_c-1},
\end{aligned}
$$
(5.3.6)

where d_j is the jth symbol of the symbol set Ω_D and $d(0)$ represents the first symbol participating in the evaluation of the CM value, without loss of generality. The conditional probability is defined in similar manner for all possible values of d. Note, that

the probability distribution of the conditional CM value $p(z|d)$ has the same probability space as an CM defined over $N_c - 1$ symbols, but it has different CM value space Ω_Z.

Given that we know the probability mass functions of the transmitted symbols and the CM component, we can calculate the conditional mutual information. Note, that this is a straightforward way of achieving the lower bound for the DDST receiver BER performance, and obtaining this limit by any other means has not been presented in the literature, based on the author's best knowledge.

After we have numerically calculated the mutual information for traditional PAM transmission, DDST without apriori knowledge of the CM component and DDST with apriori knowledge of the CM component, we can obtain the noise variance for the equivalent soft bit presentation by using the conversion of mutual information to noise variance presented in [102]. Combining this idea with the Q-function, we can obtain bit error rate estimates based on the mutual information.

In [102], the log-likelihood ratio values for a binary sequence are defined as $LLR(b) = \mu_A b + n_A$, where $b \in [-1, 1]$, n_A is a Gaussian noise component with variance σ_A^2. From this model, a bitwise mutual information for Gaussian log-likelihood ratio distributions with variance $E[|n_A|^2] = \sigma_A^2$ and mean $\mu_A = \sigma_A^2/2$ is given by the *J*-function. The *J*-function is rewritten here as

$$J(\sigma_A) = 1 - \int_{-\infty}^{\infty} \frac{\exp\left[\frac{-(t-\sigma_A^2/2)^2}{2\sigma_A^2}\right]}{\sqrt{2\pi\sigma_A^2}} log_2[1 + \exp(-t)]dt. \qquad (5.3.7)$$

The *J*-function maps the given noise variance σ_A^2 into a mutual information value I_A and J^{-1} maps a given mutual information value I_A into a equivalent noise variance σ_A^2. Note, that neither J nor its inverse J^{-1} can be written in a closed form. We have used an approximative inverse of the J-function to convert the mutual information into noise variance. The approximative J-function [21], is given as

$$\tilde{J}(\sigma_A) = \left(1 - 2^{-0.3073\sigma_A^{2*0.8935}}\right)^{1.1064} = I_A, \qquad (5.3.8)$$

and the inverse of the J-function is defined as

$$\tilde{J}^{-1}(I_A) = \left[\frac{-1}{0.3037} \log_2\left(1 - I_A^{(1/1.1064)}\right)\right]^{(1/2*0.8935)} = \sigma_A. \qquad (5.3.9)$$

Instead of looking at the obtained noise variance σ_A^2 as the noise variance of the log-likelihood ratio, we can look it as noise on top of a binary signaling. Thus, we can use it to define the BER of a binary signal in an equivalent Gaussian channel. As shown in [102], we can write $\sigma_A^2 = 4/\sigma_{n_e}^2$ and the equivalent Gaussian channel model is given as $y_e = b_e + n_e$, where the binary symbol takes values $b_e \in [-1,1]$. Given the Gray mapping assumption, we can look at each layer of bits in a symbol as an equivalent Gaussian channel. Based on this definition, we can write bit error probability based on a mutual information as

$$P(bit\ error) = Q(\tilde{J}^{-1}(I(D;Y|Z)/q)/2) = Q(1/\sigma_{n_e}). \qquad (5.3.10)$$

The maximum value of the mutual information has to be normalized to one, because the transfer function $\tilde{J}^{-1}$ is defined for binary case. This result provides us the bit error probability averaged over all bit layers.

5.3.1 Additional Comments on the Conditional Mutual Information

If we look at the vector presentation of samples related to evaluation of certain cyclic mean value, one notices that the conditional MI given by (5.3.3) over this vector of samples corresponds actually to the unconditioned mutual information $I(\mathbf{D};\mathbf{Y})$. This can be shown by considering a vector model of transmission

$$\mathbf{y}_{i,N_c \times 1} = \sqrt{1-\gamma}(\mathbf{d}_{i,N_c \times 1} - \mathbf{z}_{i,N_c \times 1}) + \sqrt{\gamma}\mathbf{p}_{i,N_c \times 1} + \mathbf{n}_{i,N_c \times 1}, \qquad (5.3.11)$$

where $i = [0,1,\dots,N_p - 1]$, $\mathbf{y}_i = [y(i), y(i+N_p)\dots y(i+(N_c-1)N_p]$ is a vector of received samples, $\mathbf{d}_i = [d(i), d(i+N_p),\dots, d(i+(N_c-1)N_p]$ is a vector of transmitted symbols with mean z_i, $\mathbf{z}_i = z_i\mathbf{1}_{N_c \times 1}$ is vectorized mean component, $\mathbf{p}_i = p_i\mathbf{1}_{N_c \times 1}$ is the pilot symbol repeated on top of the data symbols, and $\mathbf{n}_i$ is the Gaussian noise component. For this model, the vector form conditional mutual information is given

in as

$$
\begin{aligned}
I(\mathbf{D};\mathbf{Y}|Z) \;&=\; E_Z[I(\mathbf{D};\mathbf{Y})|Z] \\
&=\; \sum_{z\in\Omega_Z}\sum_{\mathbf{d}\in\Omega_{\mathbf{D}}}\int_{\mathbf{y}\in\Omega_{\mathbf{Y}}} f(\mathbf{y},\mathbf{d},z)log_2\left(\frac{f(\mathbf{y},\mathbf{d},z)p(z)}{p(\mathbf{d},z)f(\mathbf{y},z)}\right)dy \\
&=\; \sum_{z\in\Omega_Z}\sum_{\mathbf{d}\in\Omega_{\mathbf{D}}}\int_{\mathbf{y}\in\Omega_{\mathbf{Y}}} f(\mathbf{y}|\mathbf{d},z)p(z|\mathbf{d})p(\mathbf{d})log_2\left(\frac{f(\mathbf{y}|\mathbf{d},z)p(z|\mathbf{d})p(\mathbf{d})p(z)}{p(z|\mathbf{d})p(\mathbf{d})f(\mathbf{y}|z)p(z)}\right)dy \\
&=\; \sum_{\mathbf{d}\in\Omega_{\mathbf{D}}}\int_{\mathbf{y}\in\Omega_{\mathbf{Y}}} f(\mathbf{y}|\mathbf{d})p(\mathbf{d})log_2\left(\frac{f(\mathbf{y}|\mathbf{d})p(\mathbf{d})}{p(\mathbf{d})f(\mathbf{y}|z=1/N_c\sum_{k=0}^{N_c-1}d(i+kN_p))}\right)dy \\
&\leq\; I(\mathbf{D};\mathbf{Y}),
\end{aligned}
$$

$$(5.3.12)$$

for which the fourth line follows from the conditional probability

$$
p(z|\mathbf{d}) = \begin{cases} 1, & \text{when } 1/N_c\sum_{k=0}^{N_c-1}d(i+kN_p) = z \\ 0, & \text{otherwise.} \end{cases}
$$

$$(5.3.13)$$

Now, if the data symbols are mutually independent, the final result can be simplified to $I(\mathbf{D};\mathbf{Y}) = N_c I(D;Y)$, given that $y = d\sqrt{1-\gamma}+n$. This result shows that the mutual information is upper bounded by $N_c I(D;Y)$ and thus the lower bound on the error performance with DDST is equal to basic PAM transmission with equal power allocation. The results is intuitive and follows Shannon's celebrated results [96], by stating that the mutual information upper bound for DDST symbols is the same as for traditional PAM symbols given that they have equal total power and that we have apriori knowledge of the cyclic mean component. Similar results were obtained in [17], where the analysis was performed based on vector space representation, but with one significant difference. Contradicting the results of [17], we claim that because the MI upper bound for DDST symbols is the same as for TDMT and based on our simulations the bound is tight (performance is very close to the bound), we can achieve higher throughput with DDST than with TDMT in high signal to noise ratio (SNR) conditions or when we have a high number of cyclic copies present in the cyclic mean calculations if we assume that the total packet length is limited. This implies that DDST system may carry more user data symbols, when compared to TDMT, if the receiver incorporates apriori information of the cyclic mean distribution in the detection process.
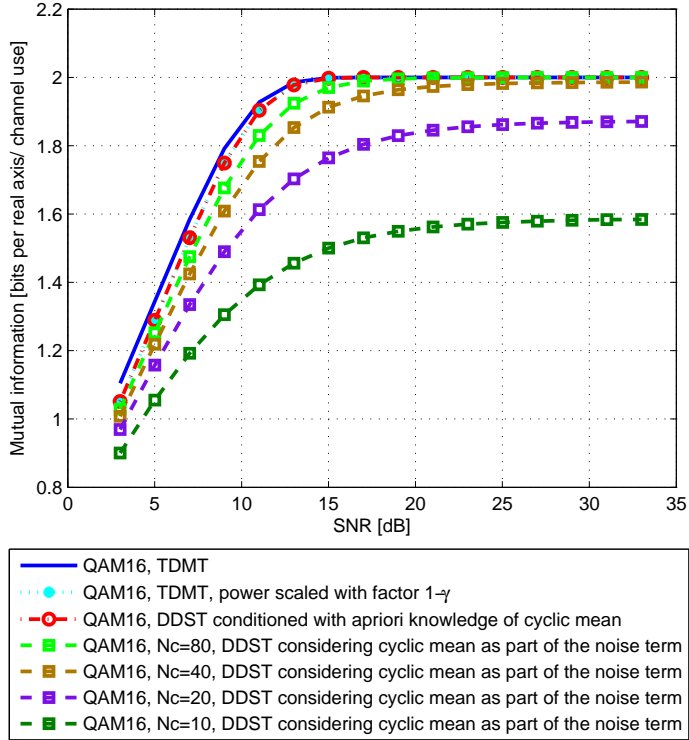
Fig. 7. Numerical MI for bits on real or imaginary axis for traditional 16-QAM constella-
tion (TDMT) and for power scaled TDMT, 16-QAM DDST with apriori knowledge
of the cyclic mean component based on (5.3.3) and 16-QAM DDST without apriori
knowledge of the cyclic mean component based on (5.3.2), with different N_c values.

5.4   Numerical Results

In this section, we compare the analytical results with simulated results. In all of the
MI results presented in this chapter, we have included the Shannon limit [96] for the
Gaussian input signals in the figures.

In Fig. 7, we show the obtained mutual information, for DDST with perfect apri-
ori information obtained by inserting (5.3.4) and (5.3.5) into (5.3.3), and without
apriori information (5.3.2) on the cyclic mean component for 16-QAM modulation.
In addition, we show the MI for power scaled TDMT in a AWGN channel based on
(5.3.12).The used signal-to-noise ratio (SNR) corresponds in our case to $SNR = 1/\sigma_n^2$,
because we have assumed that the overall transmitted signal (including $\mathbf{p}_c$) has unity

average power. This figure confirms also the discussion of Section 5.3.1, i.e., the MI of the conditional mutual information given in (5.3.3) and the mutual information for the power scaled QAM symbols (5.3.12), are the same.

As expected, the MI in the case that DDST does have apriori knowledge of the CM provides higher MI than the case that the receiver considers the CM component as interference. Interesting to notice is how quickly the MI of the receiver without apriori information increases as the number of cyclic copies increases, as is seen for the cases $N_c = 20$ and $N_c = 40$. From Fig. 7 we notice that with $N_c = 80$, with or without apriori knowledge in the receiver, the MI is quite the same. These results also depend on the used constellation, but similar results are obtained with different constellations.

Now that we have obtained the MI results for DDST, it is interesting to see what kind of BER can we expect from our system. Because we are now considering 16-QAM modulation, the number of bits per symbol in the real (imaginary) axis is $q = 2$. Based on (5.3.10), the analytic BER for receiver using apriori knowledge on the CM component is given as

$$BER_{I(\tilde{X},Y|\tilde{Z})} = Q(\tilde{J}^{-1}(I(\tilde{X};Y|\tilde{Z})/2)/2), \tag{5.4.1}$$

and the analytic BER estimate for the receiver considering the CM component as interference is given as

$$BER_{I(\tilde{X},Y)} = Q(\tilde{J}^{-1}(I(\tilde{X};Y)/2)/2). \tag{5.4.2}$$

As we can see from the Fig. 8, the analytic BER estimates are rather accurate. The hard symbol based CM estimation works with all values of cyclic copies and with $N_c = 80$, the performance is very close the lower bound. Because the CM component is the mean value of random QAM symbols, which have expected value equal to zero, it is clear that the probability of having zero valued (or very small value) for the CM component approaches unity as the number of symbols participating in the evaluation of the mean value tends to infinity. These results also indicate, that when we have relatively small number of cyclic copies, the interference induced by the CM component and its estimation requires more sophisticated methods.

The analytical tools presented in this chapter allow us to define upper and lower bounds for DDST communications system without extensive simulation campaigns. The final open question is to define the analytical solution for SER/BER with respect
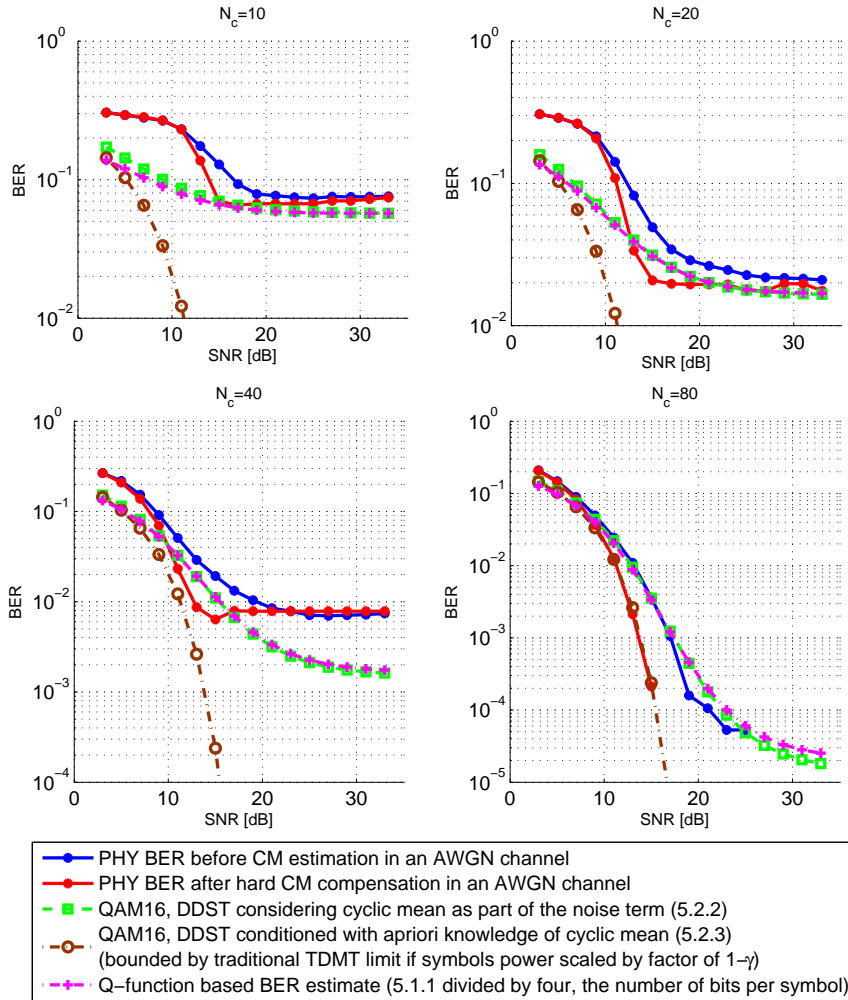
Fig. 8. *Numerical BER results, including simulated BER without cyclic mean estimation and with cyclic mean estimation, analytical BER for 16-QAM DDST with apriori (5.4.1) and without apriori information (5.4.2) of the CM component based on the MI results, with different N_c values.*

to pilot power allocation factor γ, packet lengt N_{tot}, and number of cyclic copies N_c, and to analytically compare the performance of DDST with TDMT.

6. PAPR MODELING

The peak-to-average-power-ratio (PAPR) is an important measure for mobile devices. The higher PAPR the transmitted signal has, the more linear power amplifiers (PAs) or higher power backoff has to be used to fulfill the desired spectral masks. Having better PAs means higher costs and higher power backoff means reduced link budget, which then dictates itself as shorter maximum link distances and shorter operation times in battery operated devices. The PAPR for signal x is defined as

$$PAPR(x) = 10 \log_{10} \left(\frac{\max_{x \in X}(|x|^2)}{E_X(|x|^2)} \right).$$ (6.0.3)

The traditional TDMT based SC transmission has the best PAPR properties, when compared to DDST based SC and OFDM transmissions. Inside certain transmission scheme, the PAPR depends on the used constellation. Constant amplitude constellations like binary phase shift keying (BPSK) or quadrature phase shift keying (QPSK) generate smaller PAPR than for example 256-QAM constellation. The PAPR of these constellations is easy to evaluate, but the problem becomes more complex when we include the transmitter pulse shape filter to the equations. We have considered only root-raised cosine (RRC) pulse shaping filters, but the derivations shown in [P4] hold for any other linear filter. The number of combinations affecting PAPR increases exponentially with the length of the RRC filter, so typical approach is to simulate and collect PAPR statistics and present the results with complementary cumulative distribution function (CCDF). Example distributions for SC transmission with TDMT, SIT and DDST, using pilot power allocation factor $\gamma = 0.1$ for DDST, 32 samples long RRC filter with rolloff factor $\rho = 0.2$, with QPSK, 16-QAM, and 64-QAM modulations are given in Fig. 9. We can see how the PAPR increases as we increase the constellation size. The minimum PAPR is achieved with TDMT and QPSK modulation, because QPSK is a constant amplitude modulation. The RRC filter parameters also affect the performance in all schemes. With the given parameters, at 10% PAPR probability (90% of the detected PAPR values are smaller than the depicted value in the x-axis) SIT and DDST lose 1 dB when compared to TDMT with QPSK modula-
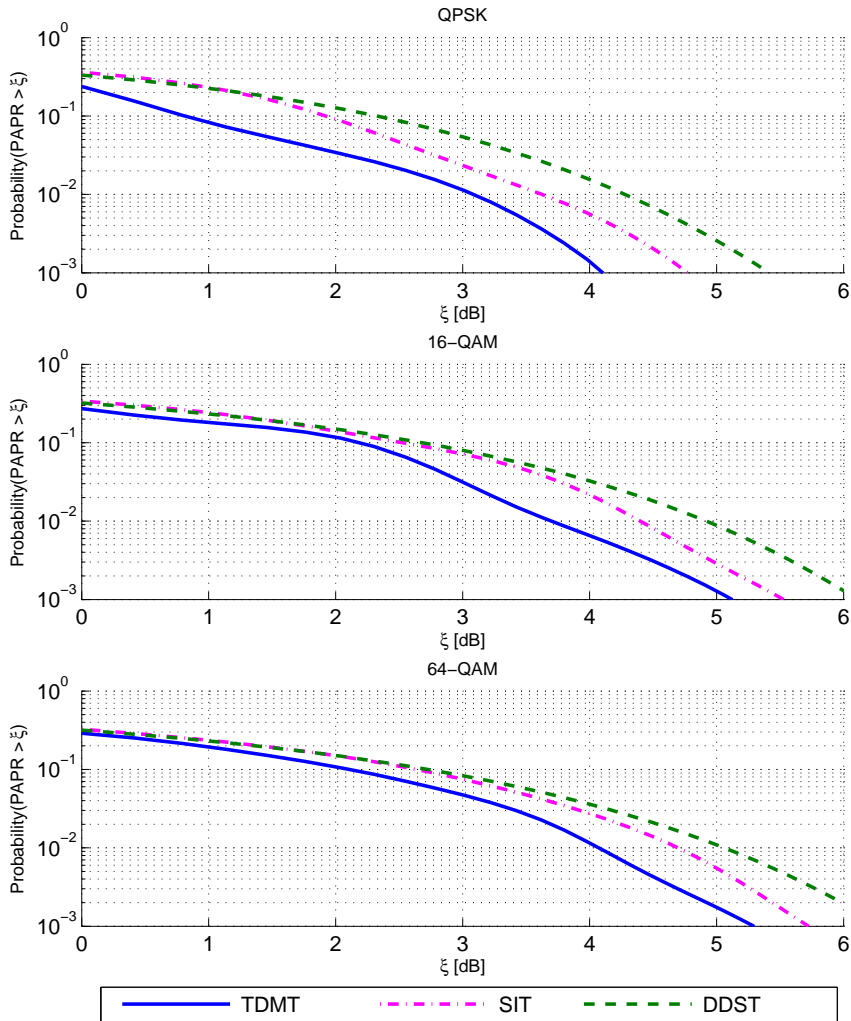
Fig. 9. *Example of the PAPR CCDF distributions for TDMT, SIT and DDST training methods when using QPSK, 16-QAM, or 64-QAM modulations. The RRC filter is 32 samples long and has a roll-off factor of 0.2. The used pilot power allocation is* $\gamma = 0.1$.

tion, but with 16-QAM and 64-QAM modulations they lose approximately 0.5 dB to TDMT. At 1% PAPR probability, the SIT is typically 0.5 dB worse than TDMT and DDST is 1 dB worse than TDMT.

If this PAPR increase of SIT or DDST is not taken into consideration, the transmitter may violate the spectral mask defined for the communications in the desired

channel. In [P4], we studied the use of symbol level amplitude limiter to minimize the PAPR increase with DDST and also proposed algorithms to improve the receiver performance.

6.1   Rapp Power Amplifier Model

The nonlinear power amplifier model we have studied is a widely-used basic model, based on solid-state power amplifier (SSPA) model by Rapp [91]. The amplitude-to-amplitude (AM-to-AM) conversion function for an input amplitude A is given as

$$G(A) = v \frac{A}{\left(1 + \left[\frac{vA}{A_0}\right]^{2p}\right)^{-2p}},$$ (6.1.1)

where v is the small signal amplification, A_0 is the saturation amplitude of the amplifier and p defines the smoothness of the transition from linear region to the limiter region.

Based on Bussgang's theorem [22], we can model the output of the power amplifier as $G(\mathbf{x}) = \alpha\sqrt{P_{AVG}}\mathbf{x} + \mathbf{n}_G$, where α is a scaling factor for the input signal, P_{AVG} is the average power of the transmitted frame, and $\mathbf{n}_G$ is uncorrelated Gaussian noise vector caused by the nonlinear power amplifier $G(\cdot)$. P_{AVG} is used to scale the average power of the transmitted frame in order to stay inside the spectral mask. The Bussgang's theorem is based on Gaussian variables, but it's results are widely used, e.g., in PAPR modeling for OFDM systems. For SC modulation, the signals are not Gaussian, but after the RRC pulse shape filter they are Gaussian like and we can apply Bussgang's theorem to model the complex phenomena caused by the power amplifier model.

One drawback with DDST in SC transmission is the increased peak power and PAPR in the transmitted signal and spectral leakage caused by the non-linear amplifier due to the increased PAPR. These problems are well known but have received relatively little attention in the recent literature. In MC transmission, the PAPR is a well known problem and the PAPR reduction has attained considerable amount of research effort during the last decade. Good overview articles in the topic are for example [49, 60] and references therein. In [25] the effect of the SIT in OFDM transmission is analyzed and upper and analytical CCDFs for peak-to-average-ratio with SIT are provided. It is shown that with both, SIT and traditional pilot based methods benefit from constant amplitude training sequences in the case of OFDM transmission. In a

SC transmission, the PAPR of the transmitted sequence is defined after the Tx pulse-shape filter. The peak power we see in the filter output depends on the maximum amplitude of the input symbols and on a portion of the absolute values of the filter coefficients, depending on the oversampling. Because we have fixed the Tx pulse-shape filter, only the maximum amplitudes of the input symbols effect the observed PAPR.

There are two main reasons for increased symbol level amplitude in DDST. First of all, we increase the amplitude range related to a certain constellation by adding a power scaled pilot sequence on top of a power scaled symbol sequence. The second main reason for increased amplitude is the possibility of a cyclic mean (data dependent pilot) component with relatively high amplitude. When this component is added on top of data and known pilot symbols, and if the angles of these complex variables happen to align, then the total symbol amplitude is significantly increased.

As discussed in [P4], the energy efficiency of the DDST transmission can be significantly improved with the use of predistortion of the transmitted signal. Several other methods exists for PAPR reduction as discussed in [49, 60] which are also applicable for SC transmission. Our target was to generate a simple predistortion scheme with iterative receiver solution achieving improved energy efficiency, especially in constant amplitude modulation cases (like QPSK modulation). As mentioned, the PAPR in SC communications has not been considered as a problem, but for example the new millimeter-wave communications [92] expected to achieve the 5G peak throughput targets [11] is expected to utilize SC communications [48] in the multi-GHz bandwidths. Achieving energy efficient power amplifiers for these new carrier frequencies and ultra-wide bandwidths and especially supporting higher modulations requires researchers to consider also PAPR reduction schemes for SC transmissions.

In [P5] the energy efficiency of SC MIMO transmission between DDST and TDMT was considered and we pointed out areas of improvement when using DDST. Unfortunately, power amplifier was not included in these simulations. We believe, based on the results of [P4], that the gains of DDST when compared to TDMT will prevail also when power amplifier nonlinearities are included in the studies, especially if the PAPR increase with DDST is handled by some method, e.g., the one presented in [P4].

In both, [P4] and [P5], the largest throughput gains for the benefit of DDST were observed with robust modulations (like QPSK) and receiver antenna diversity. This implies that DDST might also be a good solution for the ultra-reliable communications required from the future 5G systems [11].

7. SUMMARY

In this thesis, we have studied SIT, DDST, and TDMT from channel estimation perspective and DDST and TDMT from spectrally efficient data transmission perspective. The studies were initiated by defining new MSE bounds for SIT with short channel estimators [P1],[P2]. After studying SIT based channel estimators for a while, it soon became evident that DDST based system would, in our opinion, be superior. This is because the residual interference caused by self interference in SIT based channel estimation has significant impact on the channel estimation process and achieving similar throughput performance as with DDST of TDMT is very difficult. Instead, DDST does not suffer from the self interference in channel estimation, thus providing similar channel estimation accuracy as TDMT without iterative processing.

As our focus concentrated on the throughput performance comparison between DDST and TDMT, we implemented iterative, coded soft symbol based detection loop for DDST and studied the performance limits. At the same time we noted that the PAPR increase with DDST was significant, especially with QPSK modulation. Therefore, in [P3],[P4], the throughput performance with DDST and TDMT taking PAPR into consideration was studied. In [P3] the idea of limiting the amplitudes of the transmitted symbols including pilot signal, data dependent pilot signal and data signal was presented and some initial performance results were shown. The results of [P4] provide more detailed analysis and comparison between DDST and TDMT. There it is shown that with symbol level signal processing we can decrease the PAPR with DDST, especially when using QPSK, and improve the performance significantly.

Finally, in [P5], the spectral efficiency comparison between DDST and TDMT is performed in MIMO transmission scheme with variety of antenna configurations. It was shown that in cases where we have receiver antenna diversity and/or strong channel coding, DDST may provide better spectral efficiency with given parametrization. These results are open to depute because it is still only one parameter set, where especially channel length, frame length and number of estimated spatial streams effect the outcome.

Even though we were unable to analytically prove that either DDST or TDMT would be better always or in predefined certain scenarios, we have shown that with iterative signal processing in the receiver we can achieve better throughput with DDST than with TDMT, in certain scenarios. The study on DDST like methods for improving the physical layer spectral efficiency continues as we aim for higher and higher spectral efficiencies. Furthermore, combining DDST with recent achievements of sparse channel estimation [34] could provide improved performance due to reduced training overhead and comparison between TDMT using sparse sampling would be very interesting future task. In addition, we have not considered any hybrid automatic repeat and request (HARQ) coding schemes with DDST. This area could provide open topics for further research.

If DDST would lose to TDMT in the analytical solution, it may still provide additional value in certain systems or tasks. For example, when considering system synchronization, detecting ongoing transmissions from always present, known synchronization signal may be easier and sometimes even faster than detection of a scarcely transmitted synchronization pulse/packet.

Also, the possibilities of DDST like ideology with OFDM transmission has not been extensively studied, especially in the terms of throughput and always present pilot, that could also be used for synchronization. Some ideas, like the symbol-blanking scheme proposed in [44], or unitary precoding matrix based symbol spreading as proposed in [40] have been presented, but we believe that this topic could provide more if studied further. This topic has been studied also in [24, 26, 36, 69, 84, 85]. In addition, SIT or DDST combined with the new waveforms proposed for 5G communications, like the ones proposed in [15, 109, 118] or optimizing them to provide per subchannel channel estimates for FB based multiuser SC or MC communications are open topics for research. Also, recently there has been research activity in relay networks utilizing SIT, as presented in [4, 124] and references therein, which provides an open research area for optimizing training in different links and maybe even proposals for hybrid schemes using both DDST and TDMT in different phases of the transmission.

Then there are ideas of using DDST like signaling, always present and providing a couple bits of system level information, indicating for example the traffic level of a cell, or using DDST like methods for encrypting the transmitted signal. If the pilot signal would vary in time with a pseudorandom pattern known by the transmitter and receiver, the actual signal detection by an outsider would be very difficult.

One open topic is a computationally efficient, close to optimal detector structure for DDST. The per symbol probability evaluation over the cyclic mean distribution

and possible symbol values over $N_c - 1$ other symbols is computationally exhaustive search. Iterative algorithm to explore the space and iteratively updating the symbol and cyclic mean component probabilities could significantly improve the detection performance and is open for further studies. One promising area of research related to this topic could be the quantum search algorithms [20] applied to the detection of both, the data symbols and the cyclic mean component, in the receiver with reduced complexity.

APPENDIX A: COMMENTS ON MULTIPLEXING

The different multiplexing schemes in this thesis for the training signal are the time domain multiplexed training (TDMT), superimposed training (SIT) and data-dependent superimposed training (DDST). In this appendix we introduce the basic differences between these schemes and illustrate them through a simple example. In addition, short comment on the difference between single carrier (SC) and orthogonal frequency division multiplexing (OFDM) scheme is provided. For OFDM communications, a good introductory overview is given in [83].

In Fig. 10 different multiplexing schemes are shown. Sub-figure A) corresponds to TDMT where the transmitted frame starts with training sequence referred to as pilot signal and is followed by the user data signal. As shown, the signals are separated in time domain and both signals occupy the full frequency bandwidth allocated for the system. In sub-figure B), frequency domain multiplexed training is presented. This kind of approach is common especially with OFDM systems. For example, in WLAN transmission [57] the pilot carriers used for phase tracking are multiplexed like this. Most importantly, as shown in [74], this corresponds also to the transmitted signal when using DDST. Firstly, due to removing the cyclic mean from the SC signal with cycle length N_p, we obtain null carriers in the frequency domain with interval equal to the number of copies N_c. Secondly, because we have a cyclic training sequence of length $N_d = N_c * N_p$ consisting of N_c copies of the basis pilot sequence of length N_p, it has nonzero frequency bins only with separation N_c in the frequency domain. Therefore, with DDST we actually end up with frequency multiplexed training leading to interference free training signal. Removal of the frequency bins from the data signal causes the self distortion term which can be more easily approximated from the time domain signal having a known structure, like QAM modulated signal. For OFDM, direct utilization of DDST ideology would destroy user data symbols on those carriers which are nulled. Therefore, solutions like the symbol-blanking scheme [44] or unitary precoding matrix based symbol spreading [40] have been proposed to alleviate the usage of DDST also with OFDM like waveforms.

In Fig. 10, sub-figure C) corresponds to traditional SIT where the data and training
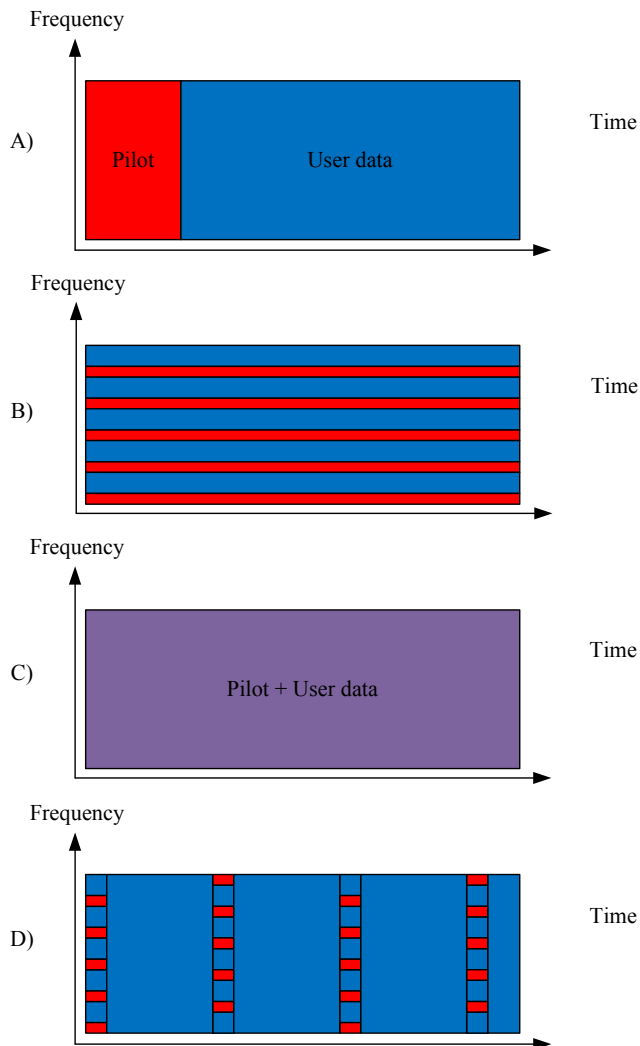
Fig. 10. *Comparison between different multiplexing schemes.*

signals are arithmetically added and there is no clear separation in either time or frequency domain. The sub-figure D) illustrates a more common approach on training symbol allocation by using both, time and frequency multiplexing, as used for example in LTE [3]. This kind of approach typically requires time and frequency domain interpolation to achieve good channel estimation performance over all subcarriers.

In Fig. 11, a simplified frequency domain illustration for SC transmission (sub-figure C)) and OFDM transmission (sub-figure B)) is shown. The sub-figure A) illustrates traditional frequency domain multiplexed transmission, where guard intervals
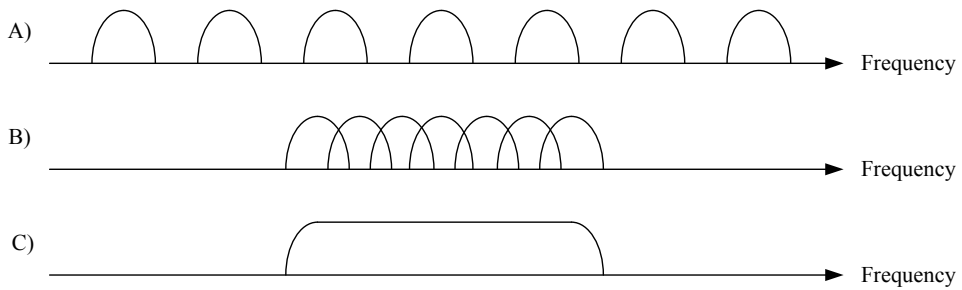
Fig. 11. *Comparison between different waveforms.*

between separated frequency channels are required. This approach is used to separate channels owned by different vendors, but the transmission inside a single channel is typically based on SC or OFDM, depending on the used communications system. Separating the allowed frequency band into channels eases the receiver design and system design. As shown in Fig. 11, OFDM is more efficient approach than basic FDM because we are capable to locate the information bearing subcarriers closer to each other.

APPENDIX B: FILTER BANK BASED STRUCTURES

Filter banks (FBs) have been successfully utilized in several different branches of signal processing, like in image processing and in audio compression. Digital communication systems heavily depend on efficient signal processing methods, so it is tempting to consider FBs also in this application [111]. FBs and FB based applications have been extensively studied at TUT. There has been research about efficient implementations of cosine or sine modulated FBs and exponentially modulated FBs [111]. Also, FB based solutions for wired transmission schemes [111] and for wireless MC communication have been studied [55, 111]. Recently, wireless SC structures with FB based FDE [122] have been researched. Efficient equalization structures are also of great interest and an important topic considering FBs and they are considered in [9, 10, 55, 93, 99, 111, 122]

The basic building blocks of a FB system are the analysis FB and the synthesis FB. These components are shown in Figures 12 and 13, respectively. In the analysis FB, a high rate input signal is bandpass filtered with M analysis filters $H_k(z)$, where $k = 0, 1, \ldots, M-1$ is the subchannel index. After the filtering, the bandpass signals are downsampled by D. If the combined sampling rate of the low rate signals is equal to the sampling rate of the high rate signal, the analysis FB is called critically sampled or maximally decimated. If the combined sampling rate is higher than the original sampling rate, we have an oversampled analysis FB.

The synthesis FB operates in a opposite way. It combines M lower rate lowpass signals to one high rate signal. Each lower rate signal is first upsampled by U and then filtered with synthesis filters $F_k(z)$, where $k = 0, 1, \ldots, M-1$ is the channel index. If the sampling rate of the high rate signal is equal to the combination of low rate sampling rates, the synthesis filter bank is called critically sampled or minimally interpolated. If the output rate is higher than the combined input rate, we have an oversampled (over-interpolated) synthesis FB, as given in [111].

Traditionally, FBs can be used in two basic configurations, analysis-synthesis and synthesis-analysis. The first one is used in signal processing tasks, where we divide the received signal into subbands, each consisting of a certain part of the received
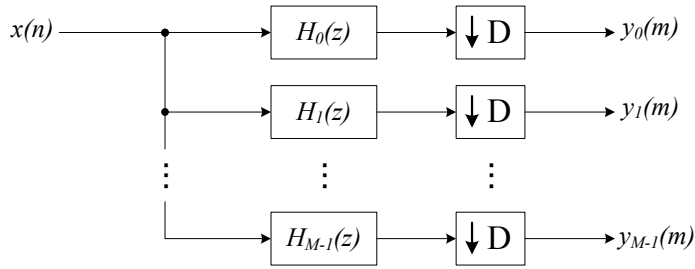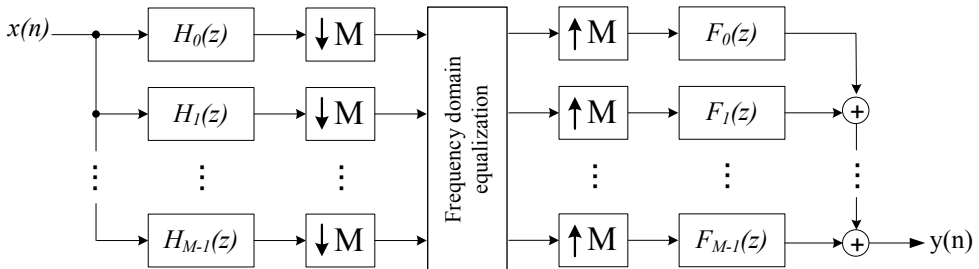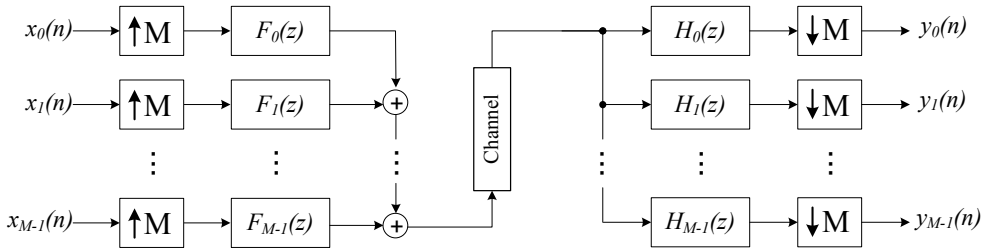
Fig. 12. Analysis filter bank.
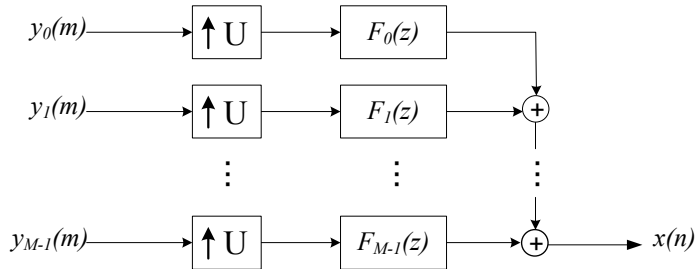
Fig. 13. Synthesis filter bank.

spectrum. These subbands are then processed independently, see example Fig. 14, where the FDE is done subband-wise. The gain from processing each subband separately comes from the lower rate of each subband and from the fact that we now have less samples to process.

Synthesis-analysis structures are used as transmultiplexer (TMUX) structures in MC systems, as shown in Fig. 15. The channel is usually omitted when designing the filter banks and the channel equalization is considered as a separate problem. Channel equalization in FB systems is discussed in more detail in Appendix C. In the MC structure, the lower rate signals are often called subchannels, referring to the

Fig. 14. Analysis-synthesis structure corresponding to single carrier system with subband-wise frequency domain equalization.

Fig. 15. *Synthesis-analysis structure modeling simplified filter bank based multicarrier transmission.*

MC ideology. The benefits of FB based MC systems are the better sidelobe attenuation compared to DFT based systems (e.g. OFDM) and the possibility of using less subchannels. When decreasing the number of subchannels, the subchannels become mildly frequency selective and require equalization structures which are able to cope with this problem. Therefore, we cannot anymore use just one complex multiplier as an equalizer.

The Nyquist pulse shaping principle is used for the subchannel signals, which means that the consecutive multicarrier symbol waveforms are overlapping in time [55]. This implies that the CP can not be used in FB based MC systems. The fact that we do not transmit the CP can be considered to improve the spectral efficiency of the FB based system.

FBs can be implemented using either perfect reconstruction (PR) or nearly perfect reconstruction (NPR) structures. PR means that if there is no signal processing between analysis and synthesis FB, the whole process can be modeled as a pure delay, caused by the analysis and synthesis filters. So, PR FB does not introduce any FB structure related errors to the processed signal. In NPR structures, small errors in the output signal are allowed. Often, the error caused by the NPR FB is known and is tolerated by the environment where the NPR system is used or is below the expected signal-to-noise ratio (SNR) threshold required for proper operation. Mainly NPR filter banks are used in FB based multicarrier systems because lower filter order can be used for reaching a given stopband attenuation level, compared to PR banks. PR FBs can be realized with efficient structures, such as fast extended lapped transform (ELT) or lattice structures [111]. For NPR FBs, there also exists very efficient implementations, like shown in [35].

APPENDIX C: CHANNEL EQUALIZATION IN FILTER BANKS

The channel equalization in filter bank (FB) based systems is performed sub-channel wise. We assume a $2\times$ oversampled input signal to the frequency domain equalizer (FDE), which is composed of $2M$ complex sub-channel signals. Sub-channel wise equalization (SCE) compensates the channel frequency response over the whole sub-channel bandwidth, including the passband and transition bands. The receiver RRC filter is also included in the equalization process. After the SCE, the real parts of the equalizer outputs are sufficient for synthesizing the signal. [1] Taking only the real part of the signal can here be considered as decimating by two, so we return to the original sampling rate which we had before the two times oversampled analysis FB. From now on the FB based subband-wise equalization is referred to as FBEQ.

In the simulations, we use a complex FIR based sub-channel equalization (CFIR-SCE). In [53], amplitude-phase adaptive sine-modulated/cosine-modulated filter bank equalizer for transmultiplexers (AP-ASCET) equalization method is presented. The AP-ASCET is an efficient sub-channel wise equalization structure, which first corrects the phase response of each sub-channel by using allpass filters and a phase rotator. After this, a FIR filter of certain length is used for amplitude equalization.

A very basic approach for sub-channel equalization is to use a CFIR-SCE, as presented in [123]. If we assume that we have 3 frequency points to equalize, a 3-tap CFIR filter [2],

$$E^{CFIR}(z) = c_0 z + c_1 + c_2 z^{-1}, \tag{C.1}$$

has the required degrees of freedom to equalize the channel frequency response within each sub-channel. This kind of structure is shown in Fig. 16. Notice that we calculate only the real part of the complex valued result, because only the real part is

---

[1] Here we assume the exponentially modulated filter bank structure of [111]. When using the more common OFDM/OQAM structure, real and imaginary parts of the sub-channel samples are used in an alternating manner.

[2] In practice, the filter is realized in the causal form $z^{-1}E^{CFIR}(z)$

used later. The sub-channel equalizer response depends on the number of frequency points considered within each sub-channel. Regarding the choice of the specific frequency points, the design can be greatly simplified when the choice is among the normalized frequencies $\omega = 0$, $\pm\frac{\pi}{2}$ and $\pm\pi$. At the selected points, the equalizer for two times oversampled FB is designed to take the following target values in order to perform MSE equalization of the channel [123]

$$\hat{W}_k = \frac{(C_k)^*/G_k}{|Q_k| + |Q_{(M+k_{mod(2M)})}| + \frac{\sigma_n^2}{\sigma_a^2}}, \tag{C.2}$$

where $C_k = H_k^{channel} G_k^2$, is the cascade of the channel impulse response and the two RRC filters of transmitter and receiver, G_k denotes the RRC filter response at the desired frequency bin k. The value $|Q_k| = G_k H_k (H_k)^* G_k = |C_k|^2/(G_k)^2$ is the channel gain and $|Q_k| + |Q_{M+k_{mod(2M)}}|$ describes the channel gain for the transition bands when the received spectrum is optimally folded [17]. For the passband, we can use simplified form of equation (C.2) by omitting the term $|Q_{M+k_{mod(2M)}}|$. The value σ_n^2/σ_a^2 is the inverse of the received SNR. The frequency index $k = 0, 1, \ldots, 2M - 1$ covers the entire spectrum band. It should be noted that here the equalizer coefficients implement the whole matched filter[3] together with the mean squared error (MSE) equalizer.

As three sub-channel frequency points are selected in the sub-channel equalizer design, there are total $4M$ frequency points for $2M$ sub-channels because the outermost two frequency points are the same for adjacent sub-channels. We consider the MSE equalizer response $\hat{W}_\psi$ at equally-spaced frequency points $\frac{\psi\pi}{2M}$, $\psi = 0, 1, \ldots, 4M - 1$. For notational convenience, we define these target frequency responses in terms of sub-channel index $k = 0, 1, \ldots, 2M - 1$, instead of frequency point index ψ. The kth sub-channel target response is denoted as η_{ik}, which is defined as

$$\eta_{ik} = \hat{W}_{2k+i}, \quad i = 0, 1, 2 . \tag{C.3}$$

At the low rate after decimation, these frequency points $\eta_{ok}, \eta_{1k}, \eta_{2k}$ are located for the even sub-channels at the normalized frequencies $\omega = \{0, \frac{\pi}{2}, \pi\}$ and for the odd sub-channels at the frequencies $\omega = \{-\pi, -\frac{\pi}{2}, 0\}$. Combining equations (C.1), (C.2) and (C.3), we get the following equations for the sub-channel equalizer frequency response $E^{CFIR}(e^{j\omega})$:

---

[3] Matched filter is the combination of the complex conjugate of the channel impulse response and the receiver RRC filter.
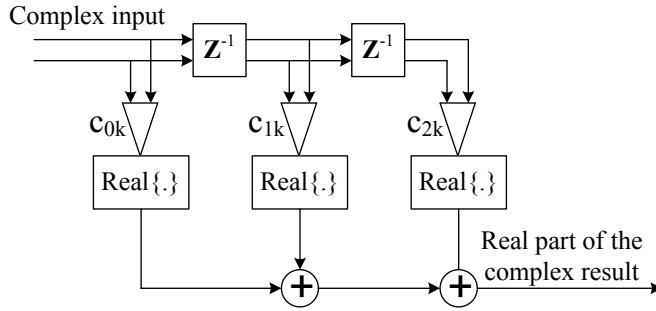
Fig. 16. *Complex FIR structure.*

$$E_k^{CFIR}(e^{j\omega}) = \begin{cases} c_{0k} + c_{1k} + c_{2k} = \eta_{0k}, & (\omega = 0) \\ jc_{0k} + c_{1k} - jc_{2k} = \eta_{1k}, & (\omega = \frac{\pi}{2}) \\ -c_{0k} + c_{1k} - c_{2k} = \eta_{2k}, & (\omega = \pi) \end{cases} \qquad \text{(C.4)}$$

for even sub-channels, and

$$E_k^{CFIR}(e^{j\omega}) = \begin{cases} -c_{0k} + c_{1k} - c_{2k} = \eta_{0k}, & (\omega = -\pi) \\ -jc_{0k} + c_{1k} + jc_{2k} = \eta_{1k}, & (\omega = -\frac{\pi}{2}) \\ c_{0k} + c_{1k} + c_{2k} = \eta_{2k}, & (\omega = 0) \end{cases} \qquad \text{(C.5)}$$

for odd sub-channels. The 3-tap complex FIR coefficients $\{c_{0k}, c_{1k}, c_{2k}\}$ of the kth sub-channel equalizer can now be obtained as follows ($+$ signs is used in even and $-$ signs in odd sub-channels):

$$\begin{aligned} c_{0k} &= \pm\frac{1}{2}\left(\frac{\eta_{0k} - \eta_{2k}}{2} - j(\eta_{1k} - \frac{\eta_{0k} + \eta_{2k}}{2})\right) \\ c_{1k} &= \frac{\eta_{0k} + \eta_{2k}}{2} \\ c_{2k} &= \pm\frac{1}{2}\left(\frac{\eta_{0k} - \eta_{2k}}{2} + j(\eta_{1k} - \frac{\eta_{0k} + \eta_{2k}}{2})\right). \end{aligned}$$

The real parts of the equalized sub-channel signals are sufficient for constructing the sample sequence for detection, so the imaginary parts are irrelevant. With this kind of reasoning it is easy to see that in the CFIR-SCE case, only two real multipliers are required to implement each of the taps. Notice, that when we take the real part of the sub-channel signal, due to the unsymmetric nature of these sub-channel signals, we

get a non-aliased mirror spectrum to the negative side with even sub-channels and to the positive side with odd sub-channels.

The number of specific frequency points used in the SCE design offer a degree of freedom and is chosen to obtain a low complexity solution. The SCE structure and the number of used points is not necessarily fixed in advance but can be determined adaptively for each sub-channel based on the frequency domain channel estimates. This enables the structure of each SCE to be controlled such that each sub-channel response is equalized optimally with the minimum number of frequency points which can be expected to result in sufficient performance. Above, solutions for a equalizer structure which utilizes three frequency points are given. We could solve these variables also for different number of frequency points as shown in [52, 123]. In the simulation results presented in this thesis, a 3-tap CFIR-SCE is used.

APPENDIX D: BROADBAND COMMUNICATIONS WITH FREQUENCY DOMAIN EQUALIZATION

Due to the increased data-rate demands, we have to utilize wider frequency bands for wireless transmission. This increases the achievable capacity of the used channel, but on the other hand makes the efficient reception of the signal more complicated. Broadband signals suffer from frequency selective fading and increased noise level compared to narrowband signals. In wideband single carrier (SC) systems, severe intersymbol interference (ISI) becomes an important issue. For these reasons, efficient equalization methods for broadband wireless transmission are continuously under intensive study. Multicarrier (MC) systems, like orthogonal frequency division multiplexing (OFDM), are seen as a solution for high data-rate mobile communications where we need efficient equalization methods. For several realized standards, e.g., wireless local area networks (WLAN), digital audio broadcasting (DAB) and digital video broadcasting (DVB), OFDM has been accepted as a good choice for transmitting data to the end users. However, MC systems have some drawbacks when considering them for mobile devices, like the need for highly linear power amplifiers [42]. By combining SC transmission with frequency domain equalization (FDE), we are able to create simpler transmitters for mobile devices and achieve similar performance as with MC systems and with similar equalization complexity [94]. For this reason SC systems with FDE are seen as interesting candidates for future transmission schemes.

Due to the increased data-rate demands, the width of the required frequency band for transmission has been increasing. The communication systems considered here are considered to be wideband systems, thus the used bandwidth is several times bigger than the coherence bandwidth of the channel. For this reason, the received signal is affected by frequency selective fading. In the time domain the reason for signal degradation is that mobile radio channels are time dispersive in nature. Time dispersion on the channel causes time spreading of the transmitted symbols and this is observed as inter symbol interference (ISI) in the received sequence. Thus, ISI is the distortion in the received symbols caused by the other transmitted symbols. The
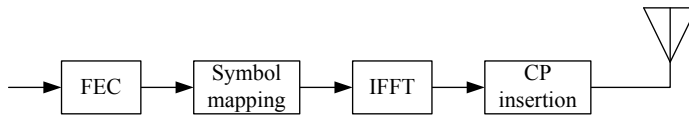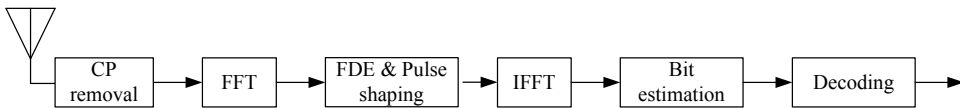
a) A simplified model of an OFDM transmitter
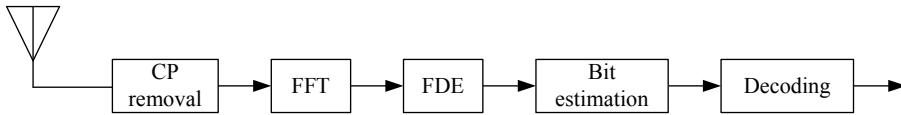
b) A simplified model of an OFDM receiver

Fig. 17. *A simplified model of an OFDM system with FDE.*
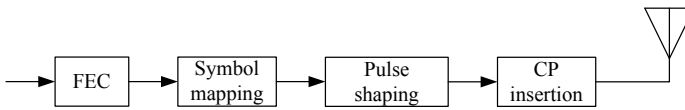
a) A simplified model of an SC transmitter

b) A simplified model of an SC receiver

Fig. 18. *A simplified model of an SC system with FDE.*

number of symbols affected by the time dispersion is linearly related to the data-rate assuming the channel is fixed.

To combat these effects, we must utilize channel equalization in the receiver. Traditionally, equalization has been done in the time domain, usually with finite impulse response (FIR) filters. One drawback of the time domain equalization is that the equalization complexity per each detected symbol grows approximately as a square of the number of symbols affected by the time dispersion [31]. Therefore, if we double the data-rate the number of symbols affected by time dispersive channel doubles and the equalization complexity in the time domain approximately quadruples.

In MC systems, the equalization is realized in the frequency domain. With FDE, the equalization complexity grows only linearly related to the number of symbols affected by time dispersion. In OFDM, each subchannel can be assumed to be flat fading with a constant phase shift. This means that each subchannel can be equalized by just one complex multiplier. In Fig. 17 a simple OFDM system model utilizing

FDE is presented. It should be noted that in OFDM, the data is transmitted block-wise and each block is extended with a cyclic prefix (CP) to remove ISI and to allow receiver to convert the linear channel convolution into a cyclic convolution which can be equalized with a simple single tap equalizer in the frequency domain.

If we move the inverse fast Fourier transform (IFFT) operation from the transmitter to the receiver after the FDE block we get an SC system with FDE, as presented in Fig. 18. The main difference between these MC and SC systems, at the conceptual level, is the location of the IFFT block. The basic structure for a SC system with FDE is presented in Fig. 18. Notice the simplified structure of the SC transmitter, where we have only *FEC*, *Symbol mapping*, *Pulse shaping* and *CP insertion* blocks. From this it is clear why the SC transmission scheme is especially interesting for the uplink communications, as most of the complexity is concentrated into the receiver (in case of uplink transmission, receiver is located in the base station). An oversampled symbol sequence is typically assumed in the SC case alleviating the realization of the receiver pulse shape filter efficiently together with the FDE. Furthermore, OFDM transmissions are also typically filtered by pulse shaping filters to reduce the out-of-band radiated power to reduce the interference between neighboring channels, although not drawn in Fig. 17.

Another method to perform the time-frequency transform is to use filter banks (FB) instead of FFT and IFFT operations. In our SC receiver model, we have an analysis FB instead of the FFT block and a synthesis FB replacing the IFFT block. The structure of these components is explained in more detail in Appendix B. The benefits of using FBs instead of Fourier transform include higher side lobe attenuation for the subbands and thus the usage of narrower guard bands in frequency domain [54]. As a drawback one can consider the higher complexity of the FB based structures.

On a radio channel, the transmitted frames interfere with each other, because of the time delayed multipath components summing up in the receiver antenna. This interference is seen in the receiver as inter block interference (IBI) in SC transmission. This interference can be efficiently reduced by using a CP in the transmitted frames [82]. The CP is a prefix that is copied from the end of the MC symbol or SC frame and added to the beginning of the symbol or frame, respectively. The length of the CP should be longer than the delay spread caused by the channel. If this is true, the received SC frames are in theory IBI free after CP removal.

The usage of the CP is an important part of both, MC and SC systems with FDE. The idea to use the CP also with a SC system was first presented in [94]. When using a CP in an SC transmission system, we can interpret that the time domain transmission of

several frames is a cyclic signal. Thus, we can model the linear convolution between the transmitted packet and a discrete channel model with a cyclic convolution which allows us to model time domain convolutions as a frequency domain multiplication between the channel response and the received packet. This greatly simplifies the equalizer design.

When considering FB based FDE, CP can not be utilized because the basis functions of the subband filters are longer than one symbol block. This can also be seen as a benefit which improves the spectral efficiency since the CP is only an undesired overhead. For OFDM, also CP free methods have been proposed based on overlapping processing e.g. in [97]. Furthermore, when we use oversampled FBs, frequency domain filtering is quite effective in suppressing strong interfering spectral components in the stopband regions of the root raised cosine (RRC) filter. RRC filters are used in transmitter and receiver as pulse shaping filters and their goal is to decrease ISI. Following from the good stopband attenuation, the receiver's FB can implement part of the receiver channel selectivity without any additional cost which indicates that we can simplify the front-end filtering and thus lower the cost-per-unit.

In [117], it was presented that by using a unique word (UW) instead of a CP in a FFT/IFFT based SC system using continuous transmission, we can use a cyclic convolution between received packet and channel response. This way, also with the UW, we can model the time domain convolution with simple multiplication in the frequency domain. The use of the UW is much more beneficial, because then also the receiver knows what has been sent, and this information can be used for synchronization and equalization purposes. Synchronization is realizable also with the CP but less can be done considering the channel estimation. Recently, in [48] a null-prefix single carrier transmission for ultra-wide communications in millimeter-wave carriers [92] was proposed. There the null sequence acts as a guard period, generates the cyclic structure to the signal, and allows time for beam switching between users or transmission directions.

One problem of the MC systems is the high PAPR. It causes severe out-of-band radiation (spectral spreading) of the signal and inband distortions to the signal, when using nonlinear power amplifiers (PA). Efficient linear PAs are difficult and expensive to realize, so this is a serious problem when considering mobile devices. There are ways to reduce the PAPR of MC systems, but they are usually either computationally expensive (e.g., adaptive constellation idea) or cause degradation in capacity (e.g., the use of dummy subcarriers). MC systems also set tight constraints on carrier synchronization and timing jitter. For example, the maximum allowed error in carrier frequency is only a fraction of the subcarrier spacing and its absolute value is

decreased as we increase the number of subcarriers within a given band. On the other hand, SC systems are robust against PAs nonlinear distortions and do not require so strict performance from carrier synchronization or timing jitter. The SC ideology is well known and has been used for decades, so it provides a well-known basis for future mobile radio systems.

BIBLIOGRAPHY

[1] 3rd Generation Parthership Project, "3GPP TS 36.211 V11.3.0 (2013-06); 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation (Release 11)," [Online], Available: http://www.3gpp.org/ftp/Specs/archive/36_series/36.211/36211-b30.zip, Accessed 20 Aug 2013.

[2] ——, "3GPP TS 45.001 V11.3.0 (2013-08); 3rd Generation Partnership Project; Technical Specification Group GSM/EDGE Radio Access Network; Physical layer on radio path; General description (Release 12)," [Online], Available: http://www.3gpp.org/ftp/Specs/archive/45_series/45.001/45001-c00.zip, Accessed 25 Oct 2013.

[3] ——, "3GPP TS 36.300 V11.7.0 (2013-09); 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2 (Release 11)," pp. 1–209, 2013.

[4] S. Abdallah and I. N. Psaromiligkos, "Semi-Blind Channel Estimation with Superimposed Training for OFDM-Based AF Two-Way Relaying," *Wireless Communications, IEEE Transactions on*, vol. 13, no. 5, pp. 2468–2467, May 2014.

[5] N. Al-Dhahir and J. Cioffi, "MMSE decision-feedback equalizers: finite-length results," *Information Theory, IEEE Transactions on*, vol. 41, no. 4, pp. 961–975, Jul 1995.

[6] E. Alameda-Hernandez, D. McLernon, M. Ghogho, A. Orozco-Lugo, and M. Lara, "Improved synchronisation for superimposed training based channel estimation," in *Statistical Signal Processing, 2005 IEEE/SP 13th Workshop on*, July 2005, pp. 1324–1329.

[7] ——, "Synchronisation for superimposed training based channel estimation," *Electronics Letters*, vol. 41, no. 9, pp. 565–567, April 2005.

[8] E. Alameda-Hernandez, D. McLernon, A. Orozco-Lugo, M. Lara, and M. Ghogho, "Frame/Training Sequence Synchronization and DC-Offset Removal for (Data-Dependent) Superimposed Training Based Channel Estimation," *Signal Processing, IEEE Transactions on*, vol. 55, no. 6, pp. 2557–2569, June 2007.

[9] J. Alhava, "Time-domain equalizer for filter bank-based multicarrier communications," in *Communications, 2001. ICC 2001. IEEE International Conference on*, vol. 1, Jun 2001, pp. 184–188 vol.1.

[10] J. Alhava and M. Renfors, "Adaptive sine-modulated/cosine-modulated filter bank equalizer for transmultiplexers," in *in Proc. European Conference on Circuit Theory and Design, ECCTD '01*, 28-31 August 2001.

[11] J. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. Soong, and J. Zhang, "What Will 5G Be?" *Selected Areas in Communications, IEEE Journal on*, vol. PP, no. 99, pp. 1–1, 2014.

[12] J. R. Barry, E. A. Lee, and D. G. Messerschmitt, *Digital Communication*, 3rd ed. Springer, 2004. [Online]. Available: http://chess.eecs.berkeley.edu/pubs/917.html

[13] C. Belfiore and J. Park, J.H., "Decision feedback equalization," *Proceedings of the IEEE*, vol. 67, no. 8, pp. 1143–1156, Aug 1979.

[14] N. Benvenuto, R. Dinis, D. Falconer, and S. Tomasin, "Single carrier modulation with nonlinear frequency domain equalization: An idea whose time has come; again," *Proceedings of the IEEE*, vol. 98, no. 1, pp. 69–96, Jan 2010.

[15] G. Berardinelli, F. M. Tavares, T. B. Sorensen, P. Mogensen, and K. Pajukoski, "Zero-tail DFT-spread-OFDM signals," in *Globecom Workshops (GC Wkshps), 2013 IEEE*, Dec 2013, pp. 229–234.

[16] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon Limit Error-Correcting Coding and Decoding: Turbo-Codes," in *IEEE International Conference on Communications*, vol. 2, Geneva, May 1993, pp. 1064–1070, conference Record.

[17] P. Bohlin and M. Coldrey, "A Comment on the Bandwidth Expansion of Data Dependent Superimposed Training," in *7th International Symposium on Wireless Communication Systems (ISWCS)*, 2010, pp. 214–217.

[18] P. Bohlin and M. Tapio, "Optimized data aided training in MIMO systems," in *Vehicular Technology Conference, 2004. VTC 2004-Spring. 2004 IEEE 59th*, vol. 2, May 2004, pp. 679–683 Vol.2.

[19] ——, "Performance evaluation of MIMO communication systems based on superimposed pilots," in *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on*, vol. 4, May 2004, pp. iv–425–iv–428 vol.4.

[20] P. Botsinis, S. X. Ng, and L. Hanzo, "Quantum search algorithms, quantum wireless, and a low-complexity maximum likelihood iterative quantum multi-user detector design," *Access, IEEE*, vol. 1, pp. 94–122, 2013.

[21] F. Brännström, L. Rasmussen, and A. Grant, "Convergence analysis and optimal scheduling for multiple concatenated codes," *Information Theory, IEEE Transactions on*, vol. 51, no. 9, pp. 3354–3364, Sept 2005.

[22] J. J. Bussgang, "Crosscorrelation functions of amplitude-distorted gaussian signals," Res. Lab. of Electronics, M.I.T., Cambridge, Mass., Tech. Rep. 216, March 26 1952.

[23] R. Carrasco-Alvarez, R. Parra-Michel, A. Orozco-Lugo, and J. Tugnait, "Enhanced channel estimation using superimposed training based on universal basis expansion," *Signal Processing, IEEE Transactions on*, vol. 57, no. 3, pp. 1217–1222, March 2009.

[24] N. Chen and G. Zhou, "What is the price paid for superimposed training in OFDM?" in *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on*, vol. 4, May 2004, pp. iv–421–iv–424 vol.4.

[25] ——, "Superimposed training for OFDM: a peak-to-average power ratio analysis," *Signal Processing, IEEE Transactions on*, vol. 54, no. 6, pp. 2277–2287, June 2006.

[26] N. Chen, G. Zhou, and A. Swami, "Bandwidth and power efficiency considerations for optimal training in OFDM," in *Statistical Signal Processing, 2005 IEEE/SP 13th Workshop on*, July 2005, pp. 1364–1369.

[27] D. Chu, "Polyphase codes with good periodic correlation properties (corresp.)," *Information Theory, IEEE Transactions on*, vol. 18, no. 4, pp. 531–532, Jul 1972.

[28] C. Ciochina and H. Sari, "A review of OFDMA and single-carrier FDMA," in *Wireless Conference (EW), 2010 European*, April 2010, pp. 706–710.

[29] J. Cioffi, G. Dudevoir, M. Eyuboglu, and G. Forney, "MMSE decision-feedback equalizers and coding. II. Coding results," *Communications, IEEE Transactions on*, vol. 43, no. 10, pp. 2595–2604, Oct 1995.

[30] J. Cioffi, G. Dudevoir, M. Vedat Eyuboglu, and J. Forney, G.D., "MMSE decision-feedback equalizers and coding. I. Equalization results," *Communications, IEEE Transactions on*, vol. 43, no. 10, pp. 2582–2594, Oct 1995.

[31] M. V. Clark, "Adaptive Frequency-Domain Equalization and Diversity Combining for Broadband Wireless Communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1385–1395, 1998.

[32] M. Coldrey and P. Bohlin, "Training-Based MIMO Systems; Part I: Performance Comparison," *Signal Processing, IEEE Transactions on*, vol. 55, no. 11, pp. 5464–5476, Nov 2007.

[33] ——, "Training-Based MIMO Systems: Part II; Improvements Using Detected Symbol Information," *Signal Processing, IEEE Transactions on*, vol. 56, no. 1, pp. 296–303, Jan 2008.

[34] S. Cotter and B. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *Communications, IEEE Transactions on*, vol. 50, no. 3, pp. 374–377, Mar 2002.

[35] F. Cruz-Roldan and M. Monteagudo-Prim, "Efficient implementation of nearly perfect reconstruction FIR cosine-modulated filterbanks," *Signal Processing, IEEE Transactions on*, vol. 52, no. 9, pp. 2661–2664, Sept 2004.

[36] T. Cui and C. Tellambura, "Superimposed pilot symbols for channel estimation in OFDM systems," in *Global Telecommunications Conference, 2005. GLOBECOM '05. IEEE*, vol. 4, Dec 2005, pp. 5 pp.–2233.

[37] R. Dinis, C.-T. Lam, and D. Falconer, "Joint Frequency-Domain Equalization and Channel Estimation Using Superimposed Pilots," in *IEEE Wireless Communications and Networking Conference, WCNC*, Las Vegas, NV, USA, March 31 - April 3 2008, pp. 447 – 452.

[38] M. Dong, L. Tong, and B. Sadler, "Optimal pilot placement for time-varying channels," in *Signal Processing Advances in Wireless Communications, 2003. SPAWC 2003. 4th IEEE Workshop on*, June 2003, pp. 219–223.

[39] ——, "Optimal insertion of pilot symbols for transmissions over time-varying flat fading channels," *Signal Processing, IEEE Transactions on*, vol. 52, no. 5, pp. 1403–1418, May 2004.

[40] G. Dou, C. He, C. Li, and J. Gao, "Channel estimation and symbol detection for OFDM systems using data-nulling superimposed pilots," *Electronics Letters*, vol. 50, no. 3, pp. 179–180, January 2014.

[41] A. Duel-Hallen and C. Heegard, "Delayed decision-feedback sequence estimation," *Communications, IEEE Transactions on*, vol. 37, no. 5, pp. 428–436, May 1989.

[42] D. Falconer, S. Ariyavisitakul, A. Benyamin-Seeyar, and B. Eidson, "Frequency domain equalization for single-carrier broadband wireless systems," *Communications Magazine, IEEE*, vol. 40, no. 4, pp. 58–66, Apr 2002.

[43] R. Gallager, "Low-density parity-check codes," *Information Theory, IRE Transactions on*, vol. 8, no. 1, pp. 21–28, January 1962.

[44] E. Gayosso-Rios, M. Lara, A. Orozco-Lugo, and D. McLernon, "Symbol-blanking superimposed training for orthogonal frequency division multiplexing systems," in *Wireless Communication Systems (ISWCS), 2010 7th International Symposium on*, 2010, pp. 204–208.

[45] M. Ghogho, "Channel and DC-Offset Estimation Using Data-Dependent Superimposed Training," in *The 2nd IEE/EURASIP Conference on DSPenabledRadio*, South Hampton, USA, 19-20 Sept. 2005, p. 5 pp.

[46] M. Ghogho, D. McLernon, E. Alameda-Hernandez, and A. Swami, "Channel estimation and symbol detection for block transmission using data-dependent superimposed training," *Signal Processing Letters, IEEE*, vol. 12, no. 3, pp. 226–229, March 2005.

[47] M. Ghogho and A. Swami, "Improved channel estimation using superimposed training," in *Signal Processing Advances in Wireless Communications, 2004 IEEE 5th Workshop on*, July 2004, pp. 110–114.

[48] A. Ghosh, T. Thomas, M. Cudak, R. Ratasuk, P. Moorut, F. Vook, T. Rappaport, G. MacCartney, S. Sun, and S. Nie, "Millimeter wave enhanced local area systems: A high data rate approach for future wireless networks," *Selected Areas in Communications, IEEE Journal on*, vol. PP, no. 99, pp. 1–1, 2014.

[49] S. H. Han and J. H. Lee, "An overview of peak-to-average power ratio reduction techniques for multicarrier transmission," *Wireless Communications, IEEE*, vol. 12, no. 2, pp. 56–65, April 2005.

[50] S. He, J. K. Tugnait, and X. Meng, "On Superimposed Training for MIMO Channel Estimation and Symbol Detection," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 3007–3021, 2007.

[51] P. Hoeher and F. Tufvesson, "Channel Estimation with Superimposed Pilot Sequence," in *Global Telecommunications Conference 1999, GLOBECOM '99*, vol. 4, 1999, pp. 2162–2166.

[52] T. Ihalainen, A. Ikhlef, J. Louveaux, and M. Renfors, "Channel Equalization for Multi-Antenna FBMC/OQAM Receivers," *Vehicular Technology, IEEE Transactions on*, vol. 60, no. 5, pp. 2070–2085, Jun 2011.

[53] T. Ihalainen, T. Stitz, and M. Renfors, "Efficient per-carrier channel equalizer for filter bank based multicarrier systems," in *Circuits and Systems, 2005. ISCAS 2005. IEEE International Symposium on*, May 2005, pp. 3175–3178 Vol. 4.

[54] T. Ihalainen, T. Stitz, A. Viholainen, and M. Renfors, "Performance comparison of LDPC-coded FBMC and CP-OFDM in beyond 3G context," in *Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium on*, May 2006, pp. 4 pp.–.

[55] T. Ihalainen, "Filter Bank Signal Processing Techniques for Dynamic Spectrum Access Communications," Ph.D. dissertation, Tampere University of Technology, Tampere, Finland, 2011, publication 956.

[56] T. Ihalainen, Y. Yang, and M. Renfors, "Filter Bank Based Frequency-Domain Equalizers with Diversity Combining," in *IEEE International Symposium on Circuits and Systems 2008, ISCAS 2008*, 18-21 May 2008, pp. 93–96.

[57] Institute of Electrical and Electronics Engineers, "IEEE Std 802.11-2012 (Revision of IEEE Std 802.11-2007); IEEE Standard for Information technology–Telecommunications and information exchange between systems Local and

metropolitan area networks–Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," pp. 1–2793, 2012.

[58] ——, "IEEE Std 802.11ad-2012 (Amendment to IEEE Std 802.11-2012, as amended by IEEE Std 802.11ae-2012 and IEEE Std 802.11aa-2012); IEEE Standard for Information technology–Telecommunications and information exchange between systems–Local and metropolitan area networks–Specific requirements-Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 3: Enhancements for Very High Throughput in the 60 GHz Band," pp. 1–628, 2012.

[59] S. A. K. Jagannatham and B. D. Rao, "Superimposed Pilot Vs. Conventional Pilots for Channel Estimation," in *Fortieth Asilomar Conference on Signals, Systems and Computers 2006, ACSSC '06*, Pacific Grove, California USA, 29 Oct. - 1 Nov., pp. 767–771.

[60] T. Jiang and Y. Wu, "An Overview: Peak-to-Average Power Ratio Reduction Techniques for OFDM Signals," *Broadcasting, IEEE Transactions on*, vol. 54, no. 2, pp. 257–268, June 2008.

[61] A. Kammoun and K. Abed-Meraim, "Performance analysis and optimal power allocation for linear receivers based on superimposed training," *EURASIP Journal on Wireless Communications and Networking*, vol. 2013, no. 1, p. 227, 2013. [Online]. Available: http://jwcn.eurasipjournals.com/content/2013/1/227

[62] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume 1: Estimation Theory*, 1993, ISBN-10: 0133457117, ISBN-13: 978-0133457117.

[63] R. Koetter, A. Singer, and M. Tüchler, "Turbo equalization," *Signal Processing Magazine, IEEE*, vol. 21, no. 1, pp. 67–80, Jan 2004.

[64] C. Laot, A. Glavieux, and J. Labat, "Turbo equalization: adaptive equalization and channel decoding jointly optimized," *Selected Areas in Communications, IEEE Journal on*, vol. 19, no. 9, pp. 1744–1752, Sep 2001.

[65] Y. Li and L. Yang, "Channel estimation and tracking using implicit training," in *Vehicular Technology Conference, 2004. VTC2004-Fall. 2004 IEEE 60th*, vol. 1, Sept 2004, pp. 72–75 Vol. 1.

[66] A. Liavas and D. Tsipouridou, "On the Performance of the Mismatched MMSE and the LS Linear Equalizers," *Signal Processing, IEEE Transactions on*, vol. 55, no. 7, pp. 3302–3311, 2007.

[67] X. Lin, J. Andrews, A. Ghosh, and R. Ratasuk, "An overview of 3GPP device-to-device proximity services," *Communications Magazine, IEEE*, vol. 52, no. 4, pp. 40–48, April 2014.

[68] O. Longoria-Gandara, R. Parra-Michel, M. Bazdresch, and A. G. Orozco-Lugo, "Iterative Mean Removal Superimposed Training for SISO and MIMO Channel Estimation," *International Journal of Digital Multimedia Broadcasting*, vol. 2008, 2008, 9 pages, Article ID 535269, doi:10.1155/2008/535269.

[69] Y. Ma, Y. Huang, X. Zhu, and N. Yi, "Pilot embedded space-time block coded ofdm in unknown multipath fading channel," in *Vehicular Technology Conference, 2004. VTC 2004-Spring. 2004 IEEE 59th*, vol. 1, May 2004, pp. 450–454 Vol.1.

[70] D. J. C. MacKay and R. Neal, "Near shannon limit performance of low density parity check codes," *Electronics Letters*, vol. 32, no. 18, pp. 1645–, Aug 1996.

[71] K. Makarychev, Y. Makarychev, A. Romashchenko, and N. Vereshchagin, "A new class of non-shannon-type inequalities for entropies," *Communications in Information and Systems*, vol. 2, pp. 147–165, 2002.

[72] F. Mazzenga, "Channel estimation and equalization for M-QAM transmission with a hidden pilot sequence," *Broadcasting, IEEE Transactions on*, vol. 46, no. 2, pp. 170–176, Jun 2000.

[73] W. McGill, "Multivariate information transmission," *Information Theory, Transactions of the IRE Professional Group on*, vol. 4, no. 4, pp. 93–111, September 1954.

[74] D. C. McLernon, A. G. Orozco-Lugo, M. M. Lara, and E. Alameda-Hernandez, "New Results for Channel Estimation via Superimposed Training," in *Second International Symposium on Communications, Control and Signal Processing, ISCCSP 2006*, Marrakech, Morocco, 13-15 March 2006.

[75] D. McLernon, E. Alameda-Hernandez, A. Orozco-Lugo, and M. Lara, "Implicitly-trained channel estimation and equalization with zero mean input data packets," in *Signal Processing and Information Technology, 2004.*

Proceedings of the Fourth IEEE International Symposium on, Dec 2004, pp. 136–139.

[76] ——, "Performance of data-dependent superimposed training without cyclic prefix," *Electronics Letters*, vol. 42, no. 10, pp. 604–606, May 2006.

[77] X. Meng and J. Tugnait, "Doubly-selective MIMO channel estimation using superimposed training," in *Sensor Array and Multichannel Signal Processing Workshop Proceedings, 2004*, July 2004, pp. 407–411.

[78] ——, "MIMO channel estimation using superimposed training," in *Communications, 2004 IEEE International Conference on*, vol. 5, June 2004, pp. 2663–2667 Vol.5.

[79] ——, "Semi-blind channel estimation and detection using superimposed training," in *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on*, vol. 4, May 2004, pp. iv–417–iv–420 vol.4.

[80] ——, "Semi-blind time-varying channel estimation using superimposed training," in *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on*, vol. 3, May 2004, pp. iii–797–800 vol.3.

[81] S. M. A. Moosvi, D. McLernon, A. Orozco-Lugo, M. Lara, and M. Ghogho, "Carrier frequency offset estimation using data-dependent superimposed training," *Communications Letters, IEEE*, vol. 12, no. 3, pp. 179–181, March 2008.

[82] B. Muquet, Z. Wang, G. Giannakis, M. de Courville, and P. Duhamel, "Cyclic prefixing or zero padding for wireless multicarrier transmissions?" *Communications, IEEE Transactions on*, vol. 50, no. 12, pp. 2136–2148, Dec 2002.

[83] R. v. Nee and R. Prasad, *OFDM for Wireless Multimedia Communications*, 1st ed. Norwood, MA, USA: Artech House, Inc., 2000.

[84] S. Ohno and G. Giannakis, "Optimal training and redundant precoding for block transmissions with application to wireless OFDM," in *Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference on*, vol. 4, 2001, pp. 2389–2392 vol.4.

[85] ——, "Optimal training and redundant precoding for block transmissions with application to wireless OFDM," *Communications, IEEE Transactions on*, vol. 50, no. 12, pp. 2113–2123, Dec 2002.

[86] A. G. Orozco-Lugo, D. C. McLernon, and M. M. Lara, "Almost Blind Channel Estimation Using Hidden Training," in *Proceedings of the 2003 10th IEEE International Conference on Electronics, Circuits and Systems, ICECS 2003*, Sharjah, United Arab Emirates, 14-17 Dec. 2003, pp. 324–327.

[87] A. Orozco-Lugo, M. Lara, and D. McLernon, "Channel estimation using implicit training," *Signal Processing, IEEE Transactions on*, vol. 52, no. 1, pp. 240–254, Jan 2004.

[88] D. H. Pham and J. Manton, "Orthogonal superimposed training on linear precoding: a new affine precoder design," in *Signal Processing Advances in Wireless Communications, 2005 IEEE 6th Workshop on*, June 2005, pp. 445–449.

[89] J. G. Proakis, *Digital Communications*, 4th ed.   McGraw-Hill International Edition, 2000, iSBN 0-07-232111-3.

[90] S. Qureshi, "Adaptive equalization," *Proceedings of the IEEE*, vol. 73, no. 9, pp. 1349–1387, Sept 1985.

[91] C. Rapp, "Effects of HPA-Nonlinearity on a 4-DPSK/OFDM-signal for a Digital Sound Broadcasting System," in *Second European Conference on Satellite Communications, ECSC-2*, Lige, Belgium, 22-24 Oct. 1991.

[92] T. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. Wong, J. Schulz, M. Samimi, and F. Gutierrez, "Millimeter Wave Mobile Communications for 5G Cellular: It Will Work!" *Access, IEEE*, vol. 1, pp. 335–349, 2013.

[93] M. Renfors, H. Xing, A. Vibolainen, and J. Rinne, "On channel equalization in filter bank based multicarrier wireless access systems," in *Vehicular Technology Conference, 1999. VTC 1999 - Fall. IEEE VTS 50th*, vol. 1, 1999, pp. 228–232 vol.1.

[94] H. Sari, G. Karam, and I. Jeanclaude, "An analysis of orthogonal frequency-division multiplexing for mobile radio applications," in *Vehicular Technology Conference, 1994 IEEE 44th*, Jun 1994, pp. 1635–1639 vol.3.

[95] H. Schoeneich and P. Hoeher, "Semi-blind pilot-layer aided channel estimation with emphasis on interleave-division multiple access systems," in *Global Telecommunications Conference, 2005. GLOBECOM '05. IEEE*, vol. 6, Dec 2005, pp. 5 pp.–3517.

[96] C. E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, July, October 1948, Reprinted version with corrections available at http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf.

[97] C. V. Sinn, J. Gotze, and M. Haardt, "Efficient data detection algorithms in single- and multi-carrier systems without the necessity of a guard period," in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, vol. 3, May 2002, pp. III–2737–III–2740.

[98] T. B. Sorensen, P. E. Mogensen, and F. Frederiksen, "Extension of the ITU Channel Models for Wideband (OFDM) Systems," in *IEEE 62nd Vehicular Technology Conference 2005, VTC-2005-Fall*, Dallas, Texas USA, 25-28 Sept. 2005, pp. 392–396.

[99] T. H. Stitz, "Filter Bank Techniques for the Physical Layer in Wireless Communications," Ph.D. dissertation, Tampere University of Technology, Tampere, Finland, 2010, publication 919.

[100] J. Talvitie, T. Levanen, and M. Renfors, "Channel estimation in time-varying flat-fading channel using superimposed pilots with interference avoidance," in *Wireless Communication Systems (ISWCS), 2010 7th International Symposium on*, Sept 2010, pp. 209–213.

[101] ——, "Channel Equalization in Narrowband Mobile Peer-to-Peer Networks using Superimposed Pilots," in *The 6th International Symposium on Wireless CommunicationsSystems, ISWCS09*, Siena-Tuscany, Italy, 7-10 Sept. 2009, pp. 526–530.

[102] S. Ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *Communications, IEEE Transactions on*, vol. 49, no. 10, pp. 1727–1737, Oct 2001.

[103] M. Tüchler, R. Koetter, and A. Singer, "Turbo equalization: principles and new results," *Communications, IEEE Transactions on*, vol. 50, no. 5, pp. 754–767, May 2002.

[104] M. Tüchler and A. Singer, "Turbo equalization: An overview," *Information Theory, IEEE Transactions on*, vol. 57, no. 2, pp. 920–952, Feb 2011.

[105] J. Tugnait and S. He, "Doubly-selective channel estimation using data-dependent superimposed training and exponential bases models," in *Informa-*

tion Sciences and Systems, 2006 40th Annual Conference on, March 2006, pp. 375–380.

[106] ——, "Multiuser/MIMO Doubly Selective Fading Channel Estimation Using Superimposed Training and Slepian Sequences," *Vehicular Technology, IEEE Transactions on*, vol. 59, no. 3, pp. 1341–1354, March 2010.

[107] J. Tugnait and W. Luo, "On channel estimation using superimposed training and first-order statistics," *Communications Letters, IEEE*, vol. 7, no. 9, pp. 413–415, Sept 2003.

[108] J. Tugnait, L. Tong, and Z. Ding, "Single-user channel estimation and equalization," *Signal Processing Magazine, IEEE*, vol. 17, no. 3, pp. 16–28, May 2000.

[109] V. Vakilian, T. Wild, F. Schaich, S. ten Brink, and J.-F. Frigon, "Universal-filtered multi-carrier technique for wireless systems beyond LTE," in *Globecom Workshops (GC Wkshps), 2013 IEEE*, Dec 2013, pp. 223–228.

[110] A. Varma, L. Andrew, C. R. N. Athaudage, and J. Manton, "Iterative algorithms for channel identification using superimposed pilots," in *Communications Theory Workshop, 2005. Proceedings. 6th Australian*, Feb 2005, pp. 195–201.

[111] A. Viholainen, "Modulated Filter Bank Design for Communications Signal Processing," Ph.D. dissertation, Tampere University of Technology, Tampere, Finland, 2004, publication 500.

[112] A. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *Information Theory, IEEE Transactions on*, vol. 13, no. 2, pp. 260–269, April 1967.

[113] A. Vosoughi and A. Scaglione, "The best training depends on the receiver architecture," in *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on*, vol. 4, May 2004, pp. iv–409–iv–412 vol.4.

[114] ——, "On the effect of channel estimation error with superimposed training upon information rates," in *Information Theory, 2004. ISIT 2004. Proceedings. International Symposium on*, 2004, pp. 316–316.

[115] ——, "Everything you always wanted to know about training: guidelines derived using the affine precoding framework and the CRB," *Signal Processing, IEEE Transactions on*, vol. 54, no. 3, pp. 940–954, March 2006.

[116] ——, "On the effect of receiver estimation error upon channel mutual information," *Signal Processing, IEEE Transactions on*, vol. 54, no. 2, pp. 459–472, Feb 2006.

[117] H. Witschnig, T. Mayer, A. Springer, A. Koppler, L. Maurer, M. Huemer, and R. Weigel, "A different look on cyclic prefix for SC/FDE," in *Personal, Indoor and Mobile Radio Communications, 2002. The 13th IEEE International Symposium on*, vol. 2, Sept 2002, pp. 824–828 vol.2.

[118] G. Wunder et al., "5GNOW: Non-Orthogonal, Asynchronous Waveforms for Future Mobile Applications," *Communications Magazine, IEEE*, vol. 52, no. 2, pp. 97–105, 2014.

[119] Y. Yang, T. Ihalainen, and M. Renfors, "Filter Bank Based Frequency Domain Equalizer in Single Carrier Modulation," in *Proc. 14th IST Mobile and Wireless Communications Summit*, Dresden, Germany, 19-23 June 2006.

[120] Y. Yang, M. Rinne, and M. Renfors, "Filter Bank Based Frequency-Domain Equalization with Noise Prediction," in *Proc. IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC '06*, Sept. 2006, pp. 1–5.

[121] ——, "Noise Predictive Turbo Equalization for a Filter Bank Based Receiver in a Single-Carrier Transmission System," in *Proc. IEEE 65th Vehicular Technology Conference*, Dublin, Ireland, Apr. 2007.

[122] Y. Yang, "Filter Bank Based Channel Equalization in Broadband Wireless Single-Carrier Systems," Ph.D. dissertation, Tampere University of Technology, Tampere, Finland, 2007, publication 677.

[123] Y. Yang, T. Ihalainen, M. Rinne, and M. Renfors, "Frequency-Domain Equalization in Single-Carrier Transmission: Filter Bank Approach," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, no. 1, p. 010438, 2007. [Online]. Available: http://asp.eurasipjournals.com/content/2007/1/010438

[124] S. Zhang, F. Gao, H. Wang, and C. Pei, "Dynamic individual channel estimation for one-way relay networks with time-multiplexed-superimposed train-

ing," *Vehicular Technology, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2014.

[125] G. Zhou and N. Chen, "Superimposed training for doubly selective channels," in *Statistical Signal Processing, 2003 IEEE Workshop on*, Sept 2003, pp. 82–85.

[126] G. Zhou, M. Viberg, and T. McKelvey, "Superimposed periodic pilots for blind channel estimation," in *Signals, Systems and Computers, 2001. Conference Record of the Thirty-Fifth Asilomar Conference on*, vol. 1, Nov 2001, pp. 653–657 vol.1.

PUBLICATIONS

Publication P1

T. Levanen, J. Talvitie, and M. Renfors, "Improved Performance Analysis for Super-imposed Pilot Based Short Channel Estimator", in *The 11th International Wrokshop on Signal Processing Advances in Wireless Communications, SPAWC 2010*, Mar-rakech, Morocco, 20-23 June 2010.

Improved Performance Analysis for Superimposed Pilot Based Short Channel Estimator

Toni Levanen, Jukka Talvitie and Markku Renfors
Department of Communications Engineering
Tampere University of Technology
P.O.Box 553, FIN-33101, Finland
Email: {firstname.lastname}@tut.fi

Abstract—In this paper we study the MSE performance of a short ML channel estimator in a discontinuous block fading channel using superimposed pilots. The earlier analytical MSE estimates that we have seen were not concerned with ideal feedback or discontinuous block-wise transmission. In addition, we are interested in the scenario where we use shorter channel estimator than the true channel length. In this paper, we present solutions for these modeling problems and obtain improved analytic MSE estimates.

Keywords: analytical MSE limits; superimposed pilots

I. Introduction

Currently, we live in the era of wireless digital communications and constantly explore for higher throughput in this challenging environment. Even though the physical layer throughput performance has increased rapidly in the past years, there often remains a significant overhead due to signalling (for system level communication) and training information (for channel estimation). In our study, we have concentrated on reducing the overhead required by the traditional training information, referred to as pilot symbols. Traditionally, the pilot symbols are placed on specified slots in time or/and in frequency domain [1]. Another way to add training information to the transmitted signal is to directly add the pilot symbols on top of the information symbols, in time or frequency domain. For this reason, these pilots are often referred to as superimposed (SI) pilots [2]. By using SI pilots, we can improve the spectral efficiency by allowing the user information to occupy the whole spectral region designed for communications. The downside is that the user information interferes greatly with the pilot sequence and that the user data symbol to interference power ratio is decreased.

To overcome this problem of self interference, in [3] a cyclic pilot sequence structure was discussed. The main idea behind the cyclic pilot structure is to allow the utilisation of cyclic mean to improve the pilot to interference power ratio (PIPR) in the estimation process. Furthermore, in the same article optimal channel independent (OCI) training sequences were derived. We have also adopted the usage of OCI training sequences in our model because of their good properties.

We have extended the model provided in [3] to our single carrier (SC) system model with filter bank (FB) based receiver structure, presented in [4]. The channel estimates are obtained in the time domain after which the sub-channel wise equalisation is performed in the frequency domain. We are using FB based receiver structure because it provides close to ideal linear equaliser performance, it has a good spectral containment properties and it is considered as a strong candidate for future wide area network communications.

In our system model the channel estimator length is smaller than the true channel length and this causes so called aliasing error in the cyclic mean calculation. The usage of short channel estimate is considered because when using cyclic pilot sequence we want to maximise the number of cycles and this leads us to compromise between cycle (estimator) length and estimation error. In addition, we can obtain complexity savings by intentionally using shorter channel estimator, if we can allow limited error floor increase in the channel estimator mean squared error (MSE) performance. In addition, we incorporate an ideal feedback for interference cancellation caused by the user data symbols and derive the estimator MSE also for this case.

This paper is organised as follows: in Section II the system model is introduced. In section III, the main concepts of the ML channel estimation are reviewed. Next, the main contributions of [3] and [5], that are utilised in this paper are reviewed in Section IV. Then we improve the MSE estimates for discontinuous block fading channel with short channel estimator in Section V and these results are tested with simulations in Section VI. Finally, in Section VII, conclusions and future topics are provided.

II. System Model

The system design originates from the uplink assumption. Thus, the complexity of the transmitting end is kept as small as possible, and most of the complexity is placed on the receiving end. We consider SC transmission because it has the benefits of lower peak to average power ratio (PAPR) and less strict frequency synchronization requirements, compared to multicarrier systems. The very simple block level design of the transmitter is given in Fig. 1. The transmitter contains only symbol mapper, pilot insertion and the transmitter pulse shape filter.
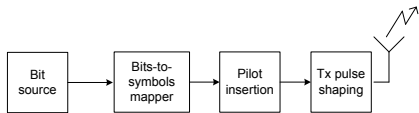
Fig. 1.   Transmitter model.

The used channel model is ITU-R Vehicular A channel with about 2.5 μs delay spread and approximately 20 MHz bandwidth [6]. The delay spread has maximum delay of 39 symbols or 78 samples in the receiver, where 2 times oversampling is utilised in the frontend.

We assume perfect synchronization in frequency and time domain, ideal down conversion and 2 times oversampling of the received signal in the Rx block, as shown in Fig. 2. Based on these normal ideality assumptions, we can present the channel between transmitter and receiver as a 2 times oversampled discrete time equivalent channel as $h(k) = |h_{Tx}(t) \otimes h_{channel}(t) \otimes h_{Rx}(t)|_{t=kT/2}$, where $\otimes$ defines a continuous-time convolution. Thus, the received symbol $z(k)$ can be given as

$$z(k) = \sum_{m=0}^{M-1} h(m)s(k-m) + w(k), \qquad (1)$$

where M is the channel length in samples, k is the time index for 2 times oversampled symbol sequence and $s(k)$ is the transmitted symbol which is zero if $k < 0$ or $k > 2L - 1$, where L is the block length in symbols. Because of the oversampling $s(k) = d(k) = p_c(k) = 0$, when k modulus 2 = 1. The noise term $w(k) = |h_{Rx}(t) \otimes v(t)|_{t=kT/2}$, where $v(t)$ is complex additive white Gaussian noise (AWGN), is simply modelled as AWGN without considering the correlation caused by the receiver pulse shape filtering with oversampling. We will see in section V that this simplification has a minor effect on the channel estimation MSE.

When we are using SI pilots, the transmitted symbols are normalised combination of user data symbols and pilot symbols, defined as

$$s(k) = \sqrt{1-\gamma}d(k) + \sqrt{\gamma}p_c(k), \qquad (2)$$

where $d(k)$ represents a data symbol, $p_c(k)$ represents a symbol from the cyclic OCI pilot sequence and γ is power normalisation factor. The power normalisation factor, γ, is used to normalise the overall transmitted symbol power to unity. This way the average transmitted power is not increased because of the SI pilots. We assume that in the transmitter the user data and pilot signal have unity power, $\sigma_d^2 = \sigma_p^2 = 1$.

From the receiver frontend, the oversampled signal is provided for channel estimator and for analysis FB. After obtaining channel estimates, sub-channel wise equalisation is performed in the frequency domain. It should be noted that the equalisation is now performed within mildly frequency selective subbands. More details on the equaliser structure can be found from [4], [7] and references there in.
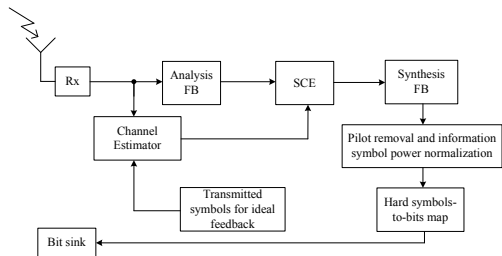

Fig. 2.   Receiver model with IF.

After sub-carrier wise equalization (SCE), the subsignals are recombined in the synthesis FB, which also efficiently realises the 2 times sampling rate down conversion. After the synthesis FB, the pilot structure is removed from the received symbol sequence, and with SI pilots the symbol sequence is power normalised as

$$\hat{d}(k) = \frac{\hat{s}(k) - \sqrt{\gamma}p_c(k)}{\sqrt{1-\gamma}}. \qquad (3)$$

After this the hard estimates are provided to the bit sink for error rate calculations and the transmitted symbols are provided for the ideal feedback (IF) loop in the receiver.

The IF loop in our receiver model implies that the correct channel response and the transmitted symbols are provided for interference cancellation (IC) operation before the cyclic mean calculation. In other words, we completely remove the interference caused by transmitted data symbols from the received data and use the distorted pilot symbol sequence for channel estimation. A more detailed block diagram of the channel estimator is provided in Fig. 3.

The IF loop provides us a lower bound on the MSE performance when using our channel estimator with hard symbol feedback. In this paper we do not consider iterative processing, but based on our practical experience we can achieve performance close to this lower bound with turbo coded system and 3 feedback iterations when using QPSK or 16-QAM modulation. For 64-QAM the performance is worse and clearly a more sophisticated system is required.

III. CHANNEL ESTIMATION

The channel estimation procedure is similar to the one presented in [3] and [5]. We find the maximum likelihood (ML) estimator for $\hat{\mathbf{h}}$, which minimises $\|\mathbf{z} - \mathbf{P_c}\hat{\mathbf{h}}\|^2$, where $\mathbf{P_c}$ is a matrix containing symbols from the cyclic pilot symbol vector $\mathbf{p}_c = [p_c(0)\ p_c(1)\ldots p_c(L-2)\ p_c(L-1)]^T$ and $\mathbf{z} = [z(0)z(1)\ldots z(L-1)]$ is the vector of received symbols.

We assume that the frame length, L, is an integer multiple of the length of one cycle, N_p, in the cyclic pilot structure $\mathbf{p}_c$. Now $L = N_cN_p$, where N_c is the number of cycles in the whole pilot vector. Thus, the cyclic pilot vector $\mathbf{p_c}$ is made of N_c copies of pilot vector $\mathbf{p}$, given as $\mathbf{p_c} = \tilde{\mathbf{I}}^T\mathbf{p}$, where $\tilde{\mathbf{I}} = [\mathbf{I}_{N_p}\ \mathbf{I}_{N_p}\ldots \mathbf{I}_{N_p}]$ is an $1 \times N_c$ block matrix and $\mathbf{I}_{N_p}$ is
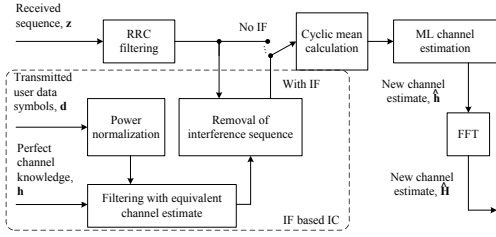
Fig. 3. SI channel estimator with interference cancellation using ideal feedback.

an $N_p \times N_p$ identity matrix. In addition, we define a matrix $\mathbf{P}$, that is a $N_p \times N_p$ cyclic matrix built from $\mathbf{p}$ as

$$\mathbf{P} = \begin{bmatrix} p(0) & p(N_p - 1) & \ldots & p(1) \\ p(1) & p(0) & \ldots & p(2) \\ \vdots & \vdots & \ddots & \vdots \\ p(N_p - 1) & p(N_p - 2) & \ldots & p(0). \end{bmatrix} \quad (4)$$

Now, we assume a continuous block wise transmission, and therefore $\mathbf{P_c} = \tilde{\mathbf{I}}^T \mathbf{P}$, which is an $N_c \times 1$ block matrix. If we also assume that the channel length M is equal to the cycle length N_p, we get [3], [5]

$$\hat{\mathbf{h}}_{without\ IF} = \frac{1}{Nc} \mathbf{P}^{-1} \tilde{\mathbf{I}}^T \mathbf{z} = \mathbf{P}^{-1} \hat{\mathbf{m}}_z, \quad (5)$$

where $\hat{\mathbf{m}}_z = [\hat{\mathbf{m}}_z(0)\ \hat{\mathbf{m}}_z(1)\ldots\hat{\mathbf{m}}_z(N_p - 1)]$, models the vector of cyclic means of the received samples, defined as

$$\hat{\mathbf{m}}_z(\iota) = \frac{1}{N_c} \sum_{\kappa=0}^{N_c-1} z(\iota + \kappa N_p). \quad (6)$$

After obtaining the channel estimate, a $4S$-point DFT of the channel estimate $\hat{\mathbf{H}} = DFT\{\hat{\mathbf{h}}\}$ is provided to the filter bank based channel equaliser. Here, S is the number of subbands in the synthesis bank, and in our simulations is set to be $S = 128$. A 3-tap complex FIR filter is used for SCE as in [7]. The equalisation structure following the channel estimator is not critical in a sense, because we are interested in the channel estimator MSE performance and not in frame or symbol error performance metrics.

IV. THEORETICAL MSE LIMITS WITH OCI TRAINING SEQUENCES

Following similar procedure as in [3] and [5], assuming OCI training sequence and by taking into account the filtered noise, the MSE of the channel estimator, $\sigma_e^2 = E\|\hat{\mathbf{h}}_{no\ feedback} - \mathbf{h}\|^2$, can be given as

$$\sigma_e^2 = \frac{(1 - \gamma)\sigma_d^2 \sum_{i=0}^{N_p-1} \sigma_{h(i)}^2 + \sigma_v^2 \sum_{l=0}^{N_{Rx}-1} \sigma_{h_{Rx}(l)}^2}{N_c \gamma \sigma_p^2}, \quad (7)$$

where $\sigma_{h(i)}^2$ is the power of the ith equivalent channel tap, $\sigma_{h_{Rx}(l)}^2$ is the power of the lth receiver RRC filter tap and

N_{Rx} is the length of the receiver RRC filter. Remember that here we have assumed that the channel length is equal to the length of one cycle in the cyclic pilot sequence.

Let us next concentrate on the IF case. If we consider the situation in the channel estimator with IF (see Fig. 3), the sampled sequence used in the channel estimation after IC is

$$\tilde{\mathbf{z}}_{IF} = \mathbf{Dh} + \mathbf{P_c}\mathbf{h} + \mathbf{w} - \mathbf{Dh} = \mathbf{P_c}\mathbf{h} + \mathbf{w}, \quad (8)$$

where $\mathbf{D}$ is a matrix of the transmitted symbols $\mathbf{d}$, and $\mathbf{w}$ is representing the filtered noise in the receiver. In the IF case, the interference caused by the user data is completely removed from the sequence. It follows that the channel estimate is then

$$\hat{\mathbf{h}}_{with\ IF} = \frac{1}{Nc} \mathbf{P}^{-1} \tilde{\mathbf{I}} \tilde{\mathbf{z}}_{IF} = \mathbf{h} + \mathbf{P}^{-1} \hat{\mathbf{m}}_w. \quad (9)$$

In this ideal case, assuming OCI pilot sequences, the MSE can be given as

$$\sigma_e^2 = E\|\hat{\mathbf{h}}_{ideal\ feedback} - \mathbf{h}\|^2 = \frac{\sigma_v^2}{N_c \gamma \sigma_p^2} \sum_{l=0}^{N_{Rx}-1} \sigma_{h_{Rx}(l)}^2, \quad (10)$$

where $E(\cdot)$ refers to a statistical expectation operator. We notice that this is the same error limit that was obtained for the data dependent superimposed training (DDST) in [5]. It is intuitive because both methods remove the interference caused by the user symbols, but in DDST the additional complexity is in the transmitting end. Also, the IF provides lower bound for the MSE with our channel estimator model for any hard feedback structure.

These error estimates are valid when the channel and the estimator have equal lengths, $N_p = M$, and when we have continuous transmission. The modeling problems arise if we consider a shorter channel estimator length than the true equivalent channel length, or when we consider discontinuous block wise transmission. In the following sections we discuss the modeling of these modifications in detail and provide analytic and simulated MSE results for comparison.

V. IMPROVED THEORETICAL MSE LIMITS WITH OCI TRAINING SEQUENCES FOR DISCONTINUOUS BLOCKWISE TRANSMISSION AND SHORT CHANNEL ESTIMATOR

Let us start with some comments on the modeling and effects of the oversampling in the receiver. Because we use two times oversampling in the receiver, we have $2L$ samples for each frame. In addition, the estimated portion of the channel has length $2N_p \leq M$. Also, the average power per sample has to be scaled by the oversampling factor, leading us to $\sigma_{d,o}^2 = \sigma_{p,o}^2 = 1/over = 0.5$ in the following equations, where $over = 2$ is the oversampling factor. We did not consider oversampling in Section III, where the earlier results were restated. To obtain oversampled versions of equations (7) and (10) we just have to scale the power terms with the oversampling factor and multiply the number of samples used in the summations with the oversampling factor.

In our considerations the channel estimator length is smaller than the true channel length and this causes error in the cyclic mean calculation. We provide an intuitive model for this error
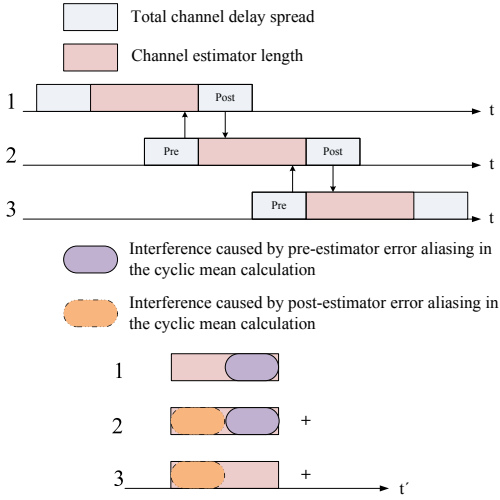
Fig. 4. Illustration of the error aliasing phenomenon when using short channel estimator. Here three copies of the cyclic pilot sequence are drawn, which have shorter length than the channel delay spread. The lower part of the figure models the summation in the cyclic mean calculation.

referred as the estimation error aliasing. This model is not accurate model of the error phenomenon, but it provides simple estimation means and relatively good estimation performance, as will be seen in the end of Section VI in Figures 6 and 7.

The basic concept of error aliasing is shown in Fig. 4. The idea is that we model the error caused by underestimating the channel response length by placing the expected channel amplitude response on top of each pilot cycle and assume that the error can be modelled as aliasing on top of the following or previous cycles. This can be thought as a simplified version of the true aliasing error caused by symbols spread in time because of the dispersive channel. As shown in Fig. 4, the pre-estimator portion of the channel response falls on top of the previous cycle and the post-estimator portion falls on top of the next cycle.

The usage of short channel estimator is considered because when using cyclic pilot sequence we want to maximise the number of cycles and this leads us to compromise between cycle (estimator) length and channel estimation error. In addition, we can obtain complexity savings by intentionally using shorter channel estimator if we can allow limited error floor increase in the channel estimator performance.

Thus, when considering discontinuous block wise transmission and short channel estimator, there are two main reason for MSE estimation inaccuracy, caused by 1) error aliasing because of shorter channel estimator length than the true channel delay spread, and 2) by modified pilot and data symbol matrices obtained in the receiver. Let us now define that the length of the true equivalent channel is $N_{channel}$, pre-estimator part of the true channel is N_{pre}, post-estimator part is N_{post} and the channel estimator length is $N_{estimator}$ where

$N_{estimator} \leq N_{channel}$ and $N_{channel} = N_{pre} + N_{estimator} + N_{post}$.

Additionally, we utilise two different MSE error metrics. First is the estimation error of the short channel estimator, which is obtained by comparing the estimate to the estimated portion of the channel, $\sigma^2_{e,short} = E\|\bar{\mathbf{h}} - \mathbf{h}_{I_{short}}\|^2$, where I_{short} corresponds to the indices related to the estimated part of the channel. This metric indicates how well the desired part of the channel is estimated. Second metric is the overall (more traditional) MSE, $\sigma^2_e = E\|\tilde{\mathbf{h}}_{extended} - \mathbf{h}\|^2$, which defines the total error between the true channel and the estimate. The difference to the first metric, $\sigma^2_{e,short}$, is equal to the sum of expected power of the channel taps outside the channel estimator. This term is defined as $\sigma^2_{e,modeling} = \sum_{i \in I_{pre,post}} E|\tilde{h}_{extended}(i) - h(i)|^2$, where $I_{pre,post}$ includes all indices not estimated by the channel estimator. Here $\hat{\mathbf{h}}_{extended}$ is the channel estimate which is extended with zeros to length $N_{channel}$. This metric can be used for symbol error rate analysis, which is one interesting future topic. Also, it is an indicator for the system designer on the compromise between improved interference cancellation through increased number of shorter cycles and increased modeling error caused by shorter channel estimate.

For ease of derivation and presentation, we assume that the length of pre-estimator part and post-estimator parts are shorter than the channel estimator length, $N_{pre} < N_{estimator}$ and $N_{post} < N_{estimator}$. The pre-estimator error is defined by $\mathbf{h}_{pre}$, where

$$\mathbf{h}_{pre} = \frac{N_c - 1}{N_c}[0 \; 0 \; \ldots \; 0 \; E|h(0)| \; E|h(1)| \; \ldots \; E|h(N_{pre}-1)|]^T. \tag{11}$$

If $N_{estimator} < N_{channel} - N_{pre}$, then we have post-estimator error aliasing, and the post-estimator error is equal to $\mathbf{h}_{post}$, where

$$\begin{aligned} \mathbf{h}_{post} = \; & \frac{N_c - 1}{N_c}[E|h(N_{pre} + N_{estimator} - 1)| \\ & E|h(N_{pre} + N_{estimator})| \ldots \\ & E|h(N_{channel} - 1)| \; 0 \; \ldots \; 0]^T. \end{aligned} \tag{12}$$

The normalisation term $(N_c - 1)/N_c$ is caused by the fact that in block-wise transmission there is no post-estimator error in the first term of the cyclic mean and there is no pre-estimator error in the last term of the cyclic mean calculation. With large number of copies this normalisation term has no significant meaning, and can be left out from the derivation. We can now define the aliasing error term $\mathbf{h}_{aliasing} = \mathbf{h}_{pre} + \mathbf{h}_{post}$, which contains pre- and post-estimator error and has length $2N_p$, which is equal to the channel estimator length in our considerations. Note that $\mathbf{h}_{aliasing}$ is now a deterministic vector.

The second error present in the simulated MSE, is caused by the approximation $\mathbf{P}_c \approx \tilde{\mathbf{I}}\mathbf{P}$. This model does not hold if we assume discontinuous block wise transmission, where there is nothing in the air before transmitting our own information block. In reality, $\mathbf{P}_c$ is made of $N_c - 1$ full copies of $\mathbf{P}$ and one copy of lower triangular matrix version of $\mathbf{P}$, referred as $\mathbf{P}_{LT}$, where everything above the main diagonal is set to zero,
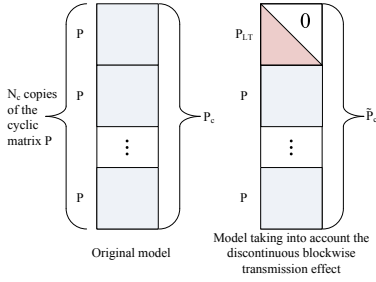
Fig. 5. Illustration of the modeling error caused by discontinuous blockwise transmission.

as shown in Fig. 5. This type of error becomes dominant as the number of cyclic copies is decreased, and is significant if only a few copies are available, e.g. less than 20 cyclic copies.

First, we have to modify the ML estimator to follow more accurately the model of discontinuous blockwise transmission. Therefore, we can easily modify the matrix $\mathbf{P}$ used in the ML estimator to incorporate the effect of the all zero portion in the $\mathbf{P}_{LT}$. Thus, the error between the approximation and model is the upper triangle part of $\mathbf{P}$, defined as

$$\mathbf{P}_{UT}(r,c) = \begin{cases} 0, & \text{if } r \geq c, \\ \mathbf{P}(r,c), & \text{otherwise.} \end{cases} \quad (13)$$

Based on this, we can generate the true ML channel estimator for the presented blockwise transmission scenario, by defining a new matrix $\tilde{\mathbf{P}} = \mathbf{P} - (1/N_c)\mathbf{P}_{UT}$, we obtain

$$\hat{\mathbf{h}} = \tilde{\mathbf{P}}^{-1}\tilde{\mathbf{m}}_z = \tilde{\mathbf{P}}^{-1}\left[\tilde{\mathbf{P}}\mathbf{h} + \mathbf{M_d}\mathbf{h} + \mathbf{m_w} + \tilde{\mathbf{P}}\mathbf{h}_{aliasing}\right]. \quad (14)$$

For this channel equaliser, we can derive the MSE limits with and without the IF in a similar manner as was done in [3] and [5]. While deriving the results, we have assumed that the products between $\mathbf{P}$ and $\mathbf{P}_{UT}$ generate diagonal matrices, which is not accurate but provides us a good approximation of the error weighting in MSE generated by the modified ML estimator structure. We obtain a weighting factor vector $\beta = \text{diag}\{\tilde{\mathbf{P}}^H\tilde{\mathbf{P}}\}$, where $\text{diag}\{\cdot\}$ generates a vector of the diagonal elements of the matrix inside the brackets. In other words, β represents the effect of the missing upper triangle portion of the first $\mathbf{P}$ matrix in $\mathbf{P}_c$ after the cyclic mean, and is defined as

$$\beta(m) = \gamma\sigma_p^2\left[N_p - \lfloor\frac{m}{over}\rfloor(2 - 1/N_c)/N_c\right], \quad (15)$$

where $m = 0, 1, \ldots, 2N_p - 1$ and $over = 2$ is the oversampling factor in the receiver frontend. For comparison, with two times oversampling, continuous transmission and with OCI sequences, $\mathbf{P}^H\mathbf{P} = \gamma\sigma_{p,o}^2 2N_p\mathbf{I}_{2N_p}$. In addition to prementioned weighting caused by missing pilots, the block wise assumption also affects the interference from user data. Clearly, if there is no earlier transmission, there is no interfering data either. This is taken into account in the MSE derivation as additional weighting factor vector, δ, of the user data related error, as

$$\delta(m,i) = \begin{cases} (1-\gamma)\sigma_{d,o}^2/N_c, & \text{if } m - i \geq 0 \\ (1-\gamma)\sigma_{d,o}^2(N_c-1)/N_c^2, & \text{otherwise,} \end{cases} \quad (16)$$

where $m, i = 0, 1, \ldots, 2N_p - 1$. Finally, for the channel estimator given in (14) we obtain MSE estimates without IF as

$$\sigma_{e,short,without\ IF}^2 = \sum_{m=0}^{2N_p-1}\frac{1}{\beta(m)}\left\{\sum_{i=0}^{2N_p-1}\delta(m,i)\sigma_{h(i)}^2 + \sigma_w^2/N_c\right\} + \|\mathbf{h}_{aliasing}\|^2, \quad (17)$$

$$\sigma_{e,without\ IF}^2 = \sum_{m=0}^{2N_p-1}\frac{1}{\beta(m)}\left\{\sum_{i=0}^{2N_p-1}\delta(m,i)\sigma_{h(i)}^2 + \sigma_w^2/N_c\right\} + \|\mathbf{h}_{aliasing}\|^2 + \sigma_{e,modeling}^2. \quad (18)$$

In similar manner, for the IF iteration we get

$$\sigma_{e,short,with\ IF}^2 = \frac{\sigma_w^2}{N_c}\sum_{m=0}^{2N_p-1}\frac{1}{\beta(m)} + \|\mathbf{h}_{aliasing}\|^2. \quad (19)$$

$$\sigma_{e,with\ IF}^2 = \frac{\sigma_w^2}{N_c}\sum_{m=0}^{2N_p-1}\frac{1}{\beta(m)} + \|\mathbf{h}_{aliasing}\|^2 + \sigma_{e,modeling}^2. \quad (20)$$

Here, $\sigma_w^2 = \sigma_v^2/over\sum_{l=0}^{N_{Rx}-1}\sigma_{h_{Rx}(l)}^2$ models the power of the receiver pulse shape filtered noise with oversampling. The analytical MSE is slightly increased due to the missing pilot symbol information in $\tilde{\mathbf{P}}$, whereas the interference caused by the user data is slightly decreased.

VI. SIMULATION RESULTS

In the presented simulations, the used constellation is 16-QAM and we used power factor $\gamma = 0.26$ to obtain the presented results. The power factor was chosen based on simulated BER results with and without IF, where the designer has to make compromise between the performance with and without IF. Higher pilot power improves the channel estimation performance but degrades the user data symbol power, which degrades the BER performance. In addition, the performance with IF always decreases if the pilot power is increased.

In Fig. 6, we have plotted the simulation based MSE values, new analytical MSE estimates without and with IF based on (17) and (19), and the old MSE estimates without and with IF given in (7) and (10). There are now $N_c = 120$ copies of a pilot sequence of length $2N_p = 64$ samples per transmitted block and the used block length is 7680 samples. The true equivalent channel length is 142 samples and it is estimated by 64 samples long estimator. Now there are 78 samples outside the estimator, from which the aliasing error can be defined. From these samples, 16 are pre-estimator and 62 are post-estimator. The new MSE estimates follow the simulated behaviour clearly better than the old ones, because of the
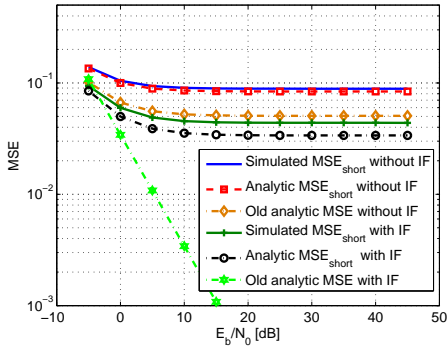
Fig. 6. Comparison between simulated and analytical MSE performance with 16-QAM constellation and 64-tap channel estimator.
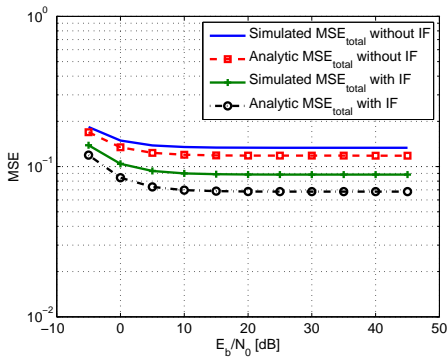

Fig. 7. Comparison between simulated and analytical total MSE performance with 16-QAM constellation and 64-tap channel estimator.

modified problem setup. Even though the error aliasing model is a simplified model, the analytic MSE estimates follow well the simulated ones.

In Fig. 7 we have plotted the simulated total MSE and the analytic MSE estimates based on (18) and (20), without and with IF, respectively. As expected, the total MSE is bigger than the MSE for the estimated portion of the channel. The analytic MSE is following the simulated values, but is slightly optimistic about the error value. Overall, the simulated values follow well the analytical ones up to $N_c = 120$, which corresponds to channel estimator length $2N_p = 64$. With channel estimator lengths shorter than this, the analytical estimates become very optimistic because of the simple error aliasing model.

VII. CONCLUSION

In this paper we have presented an interference cancelling receiver structure for SC communications with SI pilot based channel estimation and FB based channel equalisation. The interference cancelling structure is designed to shift the realisation complexity to the receiver side, e.g., to the base station. Channel equalisation is performed in frequency domain with FB based SCE, which has close to ideal linear equaliser performance.

We have restated the MSE limits for the ML channel estimator obtained in [3] and [5]. We showed that when we are interested in a discontinuous blockwise transmission in a block fading channel, we have to modify the cyclic matrix $\mathbf{P}$ to incorporate the assumption of the first transmission to the channel. In addition, we obtained a method to improve the MSE estimate in the cases when we are using a channel estimator, which is shorter than the true equivalent channel length. The presented analytical estimates follow well the simulated MSE values, and provide us important performance bounds which can be used, for example, when deriving the analytic symbol error performance of the presented receiver structure.

Furthermore, in future studies we will show that with iterative decision feedback structures using efficient channel coding we can achieve performance very close to the presented ideal feedback performance limits. This is the major motivating force behind this study. In addition, further studies will concentrate on obtaining the analytic symbol error rates for this channel estimation scheme in single antenna and multiantenna scenarios. Interesting throughput comparisons with traditional time domain multiplexed pilot symbols in single-input multiple-output channel are under preparation.

REFERENCES

[1] S. Adireddy, L. Tong, and H. Viswanathan, "Optimal Placement of Known Symbols for Frequency-Selective Block-Fading Channels," *IEEE Trans. Inf. Theory*, vol. 48, pp. 2338–2353, Aug. 2002.

[2] P. Hoeher and F. Tufvesson, "Channel Estimation with Superimposed Pilot Sequence," in *Global Telecommunications Conference 1999, GLOBECOM '99*, vol. 4, 1999, pp. 2162–2166.

[3] A. G. Orozco-Lugo, M. M. Lara, and D. C. McLernon, "Channel Estimation Using Implicit Training," *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 240–254, 2004.

[4] Y. Yang, T. Ihalainen and M. Renfors , "Filter Bank Based Frequency Domain Equalizer in Single Carrier Modulation," in *Proc. IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC '06*, Sept. 2006.

[5] D. C. McLernon et al., "New Results for Channel Estimation via Superimposed Training," in *14th European Signal Processing Conference, EUSIPCO 2006*, Florence, Italy, 4-8 Sept. 2006.

[6] T. B. Sorensen, P. E. Mogensen, and F. Frederiksen, "Extension of the ITU Channel Models for Wideband (OFDM) Systems," in *IEEE 62nd Vehicular Technology Conference 2005, VTC-2005-Fall*, Dallas, Texas USA, 25-28 Sept. 2005, pp. 392–396.

[7] Y. Yang, T. Ihalainen, M. Rinne and M. Renfors, "Frequency-domain Equalization in Single-carrier Transmission: Filter Bank Approach," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, Article ID 10438, 16 pages, 2007.

Publication P2

T. Levanen and M. Renfors, "Improved Performance Bounds for Iterative IC LMMSE Channel Estimator with SI Pilots", in *The 21st Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2010*, Istanbul, Turkey, 26-29 Sept. 2010.

Improved Performance Bounds for Iterative IC LMMSE Channel Estimator with SI Pilots

Toni Levanen and Markku Renfors
Department of Communications Engineering
Tampere University of Technology
P.O.Box 553, FIN-33101, Finland
Email: {firstname.lastname}@tut.fi

Abstract—In this paper an iterative, interference cancelling receiver structure for single carrier transmission using superimposed training is studied. We derive the analytical MSE limits for LMMSE channel estimator using perfect channel knowledge and show that the simulated values follow well the analytical ones. Then, we utilise these results to analyse the MSE performance of a combined ML-LMMSE channel estimator structure, where the initial channel estimates are obtained from the ML channel estimator.

Keywords: iterative receiver, LMMSE estimator, superimposed pilots

I. INTRODUCTION

Currently, we live in the era of wireless digital communications and constantly explore for higher throughput in this hostile environment. Even though the physical layer throughput performance has been increasing rapidly in the last years, there are, in addition, increased requirements for signalling and training information. In our study, we have concentrated on reducing the overhead required by the training information used for channel estimation, referred as pilot symbols. Traditionally the pilot symbols are placed on specified slots in time and/or in frequency [1]. Another way to add training information to the transmitted signal is to directly add the pilot symbols on top of the user data symbols. For this reason, these pilots are often referred as superimposed (SI) pilots [2]. By using SI pilots, we can improve the spectral efficiency by allowing the user information to occupy the whole spectral region designed for communications. The down side is that the user information interferes greatly with the pilot sequence and that the user data symbol power to interference power ratio is decreased.

To overcome this problem of self interference, in [3] a cyclic pilot sequence structure was discussed. The main idea behind the cyclic pilot structure is to allow the utilisation of cyclic mean to improve the pilot to interference power ratio (PIPR) in the estimation process. Furthermore, in the same article optimal channel independent (OCI) training sequences were derived which are also used in our system model. Also, the effect of DC bias was studied in [3], but it is not considered in this paper.

We have extended the model provided in [3] to our single carrier (SC) model with filter bank (FB) based receiver structure, presented in [4]. The channel estimates are obtained in time domain after which the sub-channel wise equalisation (SCE) is performed in the frequency domain. The FB based receiver structure is used because it provides close to ideal linear equaliser performance, has good spectral containment properties and is considered as a strong candidate for future wide area network (WAN) communications. The performance of the maximum likelihood (ML) receiver in the interference cancelling (IC) receiver with single-input multiple-output reception was studied in [5]. By IC we mean that we model the interference caused by user data based on latest symbol and channel response estimates, and remove this interference from the received sequence before channel estimation procedures.

In our system model the channel estimator length is smaller than the true channel length and this causes so called aliasing error in the cyclic mean calculation. The usage of short channel estimate is considered because when using cyclic pilot sequence we want to maximise the number of cycles and this leads us to compromise between cycle (estimator) length and estimation error. In addition, we can obtain complexity savings by intentionally using shorter channel estimator if we can allow limited error floor increase in the channel estimator mean squared error (MSE) performance.

For this paper, we have implemented the linear minimum mean squared error (LMMSE) channel estimator and derived the analytical MSE performance for two different setups. First, we derive the MSE performance limits for ideal LMMSE, where we assume that LMMSE knows the channel response. The second setup considers ML-LMMSE channel estimator, in which we obtain apriori channel estimate from ML channel estimator to be used in the LMMSE channel estimator.

This paper is organised as follows: in Section II the system model is introduced. The derivation of the analytical MSE for the ideal LMMSE channel estimator is given in subsection III-A. Next, in subsection III-B, the MSE results derived for the ideal LMMSE estimator are used together with results from [5] to estimate the MSE performance of the combined ML-LMMSE receiver with ML apriori channel estimation. In the end, in Section IV, conclusions and future topics are provided.
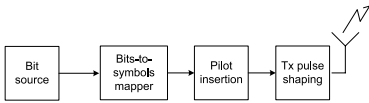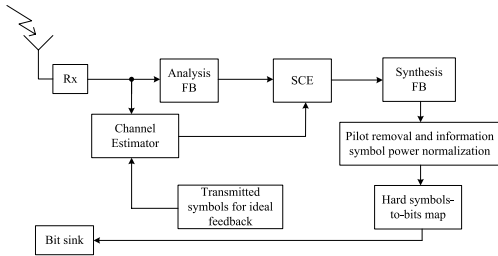
Fig. 1. Transmitter model.

Fig. 2. Receiver model for analytical MSE derivations with IC based on IF.

II. System Model

The system design originates from the uplink assumption. Thus, the complexity of the transmitting end is kept as small as possible and most of the complexity is positioned to the receiving end. The very simple block level design of the transmitter is given in Fig. 1. The transmitter contains bit source, symbol mapper, pilot insertion and the transmitter pulse shape filter.

The used channel model is ITU-R Vehicular A channel with about 2.5 μs delay spread and approximately 20 MHz bandwidth [6]. The delay spread has maximum delay of 39 symbols or 78 samples in the receiver, where 2 times oversampling is utilised in the ideal frontend.

By ideal frontend we mean that we assume perfect synchronisation in frequency and time domain, ideal down conversion and 2 times oversampling of the received signal in Rx block. The receiver model used for deriving the MSE performance is presented in Fig. 2. This model contains only one receiving antenna, but the channel estimation procedure is simply copied to each reception branch with multiple antennas.

Based on these normal ideality assumptions, we can present the channel between transmitter and receiver as a 2 times oversampled discrete time equivalent channel as $h(k) = |h_{Tx}(t) \circ h_{channel}(t) \circ h_{Rx}(t)|_{t=kT/2}$, where $\circ$ defines a continuous time convolution. The overall equivalent channel has in our case a length of 142 samples. The received symbol $z(k)$ can be given as

$$z(k) = \sum_{m=0}^{M-1} h(m)s(k-m) + w(k), \qquad (1)$$

where M is the channel length in samples, k is the time index for 2 times oversampled symbol sequence and $s(k)$ is a transmitted symbol, which is zero if $k < 0$ or $k > 2L - 1$, where L is the frame length in symbols. The noise term $w(k) = |h_{Rx}(t) \circ v(t)|_{t=kT/2}$, where $v(t)$ is complex additive

white Gaussian noise (AWGN), is simply modelled as AWGN without considering the correlation caused by the 2 times oversampled receiver pulse shape filtering. We will see in section III that this simplification has a minor effect on the channel estimation mean squared error (MSE).

When we are using SI pilots, the transmitted symbols are normalised combination of user data symbols and pilot symbols, defined as $s(k) = d(k) + p_c(k)$, where $d(k)$ represents a data symbol and $p_c(k)$ represents a symbol from the cyclic OCI pilot sequence. The transmitted user data symbol power is given as $\sigma_d^2 = 1 - \gamma$ and the pilot symbol power as $\sigma_{p_c}^2 = \gamma$, where γ is a power normalisation factor. Because of the oversampling, $s(k) = d(k) = p_c(k) = 0$ when k modulus $2 = 1$. The power normalisation factor, γ, is used so that the overall transmitted symbol power is normalised to unity, $\sigma_s^2 = 1$.

From the receiver frontend, the oversampled signal is provided for channel estimator and for analysis FB. After obtaining channel estimate, SCE is performed in the frequency domain. It should be noted that the equalisation is now performed within mildly frequency selective subbands. More details on the equaliser structure can be found from [7], [4] and references there in.

After SCE the subsignals are recombined in the synthesis FB, which also efficiently realises the 2 times sampling rate down conversion. After the synthesis FB, the pilot structure is removed from the received symbol sequence, and with SI pilots the symbol sequence power has to be normalised. Next, we have a hard symbols-to-bits mapping and the hard bit estimates are provided to the bit sink for error rate calculations. In the derivation of the MSE limits, we consider only one pass through the channel estimator without feedback, and one pass with ideal feedback (IF). Channel estimation with IF assumes that the transmitted symbols and the true equivalent channel is known by the receiver in the IC process. In other words, IF perfectly removes the interference caused by the user data symbols and provides us a lower bound for the MSE for any hard IC feedback scheme.

III. Channel Estimation

In this paper we assume that we use a short channel estimator, which has shorter length than the true channel response length. We consider this setup because with cyclic SI pilots, the channel estimation performance does not depend only on the length of the estimator but also on the number of cyclic copies per transmitted frame. Therefore, there exists an optimal compromise between the pilot cycle length and the number of cycles per frame. Of course, other parameters, like pilot power allocation factor γ affect the performance of the channel estimator, but the overall system parameter optimisation problem is greatly bigger and is not considered in this paper. Further more, we assume discontinuous block wise transmission, which leads us to modify the pilot symbol matrix present in the equations and also creates an additional weighting for the user symbol error related parameters. These phenomenons are discussed in more detail in [5].

First, we clarify the matrix notation used in this paper for analytical MSE derivations. We assume that the frame length is $L = N_c \times N_p$, where N_c is the number of cyclic copies and N_p is the length of cyclic OCI pilot sequence. One OCI pilot cycle is given as $\mathbf{p} = [p(0)\ p(1)\ \ldots\ p(N_p - 1)]^T$. The cyclic pilot sequence added on top of the transmitted frame is given as $\mathbf{p}_c = \tilde{\mathbf{I}}\mathbf{p} = [\mathbf{p}\ \mathbf{p}\ \ldots \mathbf{p}]^T$, where $\tilde{\mathbf{I}} = [\mathbf{I}\ \mathbf{I}\ \ldots \mathbf{I}]^T$ is an $N_c \times 1$ block matrix built from $N_p \times N_p$ identity matrices. The pilot matrix $\mathbf{P}_c = \tilde{\mathbf{I}}\mathbf{P} = [\mathbf{P}\ \mathbf{P}\ \ldots \mathbf{P}]^{\mathbf{T}}$ is an $N_c \times 1$ block matrix built from $N_p \times N_p$ cyclic pilot matrices $\mathbf{P}$.

Because we assume a discontinuous block wise transmission, we have to take into consideration the effect of discontinuous transmission in the pilot matrix $\mathbf{P}_c$. As it was shown in [5], we can obtain accurate estimator for discontinuous block wise transmission by defining a new matrix $\tilde{\mathbf{P}} = \mathbf{P} - (1/N_c)\mathbf{P}_{UT}$, where $\mathbf{P}_{UT}$ is an upper triangular matrix built from $\mathbf{P}$, and is defined as

$$\mathbf{P}_{UT}(r,c) = \begin{cases} 0, & \text{if } r \geq c, \\ \mathbf{P}(r,c), & \text{otherwise.} \end{cases} \quad (2)$$

We use a simple method to incorporate the additional error caused by the short channel estimator defined as error aliasing [5]. The basic idea is that the additional estimation error caused by short channel estimator is approximated by error aliasing in the cyclic mean calculations. Basically, we model the error by adding the expected amplitude response on top of each pilot cycle and study the channel tap powers aliasing on top of previous or following pilot cycles. This model causes optimistic behaviour estimates when channel estimators with length less than half of the true channel response are used. In our case, we use 96 tap channel estimator to estimate the 142 taps long true equivalent channel response. This choice of channel estimator length is arguable, because most of the equivalent channel taps outside the channel estimator reach are close to zero and are caused by the double RRC filtering in the system. Thus, the channel behavior is well estimated with our short channel estimator.

The deterministic error vector $\mathbf{h}_{aliasing} = \mathbf{h}_{pre} + \mathbf{h}_{post}$, is the aliasing error vector and consists of pre-estimator and post-estimator aliasing error vectors. The $\mathbf{h}_{pre}$ contains the expected channel amplitude response taps aliasing on top of the previous pilot cycles and the vector $\mathbf{h}_{post}$ contains the expected channel amplitude response taps aliasing on top of the following pilot cycles. This phenomenon is discussed in more detail in [5].

When defining the LMMSE channel estimator, we want to minimise the expected value of the squared error, $E\{|\mathbf{h}_{I_{short}} - \hat{\mathbf{h}}_{I_{short}}|^2\}$, where I_{short} is the set of indices which are estimated by the short channel estimator. Thus, the presented MSE estimates provide only the MSE for the estimated part of the equivalent channel response. The total MSE can be easily approximated by adding the summed power of the expected channel amplitude response outside the estimator, as was done in [5]. We define $\tilde{\mathbf{h}} = \mathbf{h}_{I_{short}}$ as the estimated part of the true equivalent channel. If we now make the assumptions that the noise and the total interference experienced by the pilot
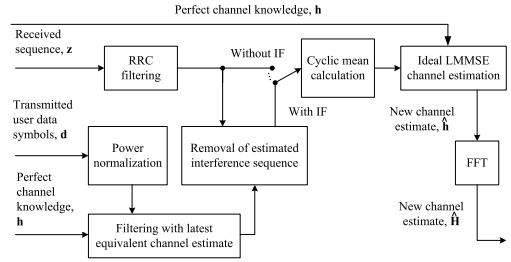

Fig. 3. Ideal LMMSE channel estimator with IC based on IF.

sequence is AWGN, channel taps are i.i.d. and have zero mean, i.e. $E\{\mathbf{h}\} = \mathbf{0}$, the LMMSE estimator can be simplified to [8]

$$\hat{\mathbf{h}} = (\sigma^2 \mathbf{C}_{\tilde{\mathbf{h}}}^{-1} + \tilde{\mathbf{P}}_c^H \tilde{\mathbf{P}}_c)^{-1} \tilde{\mathbf{P}}_c^H \mathbf{z}, \quad (3)$$

where $(\cdot)^H$, refers to a Hermitian transpose operation, and

$$\sigma^2 = \begin{cases} \sigma_w^2 + \|\tilde{\mathbf{h}}\|^2 \sigma_d^2 \text{ , without IF and} \\ \sigma_w^2 \text{ , with IF} \end{cases} \quad (4)$$

is the total interference power experienced by the pilot sequence. The channel covariance matrix $\mathbf{C}_{\tilde{\mathbf{h}}}$, contains the apriori information of the channel tap values. By the assumption of independent tap coefficients, it becomes diagonal, i.e., $\mathbf{C}_{\tilde{\mathbf{h}}} = diag\{|\tilde{h}(0)|^2, |\tilde{h}(1)|^2, \cdots, |\tilde{h}(over N_p - 1)|^2\}$. By assuming the cyclic OCI training sequence, the LMMSE estimator can be reduced to

$$\hat{\mathbf{h}} = (\frac{\sigma^2}{N_c} \mathbf{C}_{\tilde{\mathbf{h}}}^{-1} + \tilde{\mathbf{P}}^H \tilde{\mathbf{P}})^{-1} \tilde{\mathbf{P}}^H \hat{\mathbf{m}}_z. \quad (5)$$

Based on [5], we can approximate the matrix product term $\tilde{\mathbf{P}}^H \tilde{\mathbf{P}}$, concerning the new pilot matrix $\tilde{\mathbf{P}}$, as

$$\tilde{\mathbf{P}}^H \tilde{\mathbf{P}} \approx \text{diag}\left\{\sigma_p^2 \left[N_p - \lfloor \frac{i}{over} \rfloor (2 - 1/N_c)/N_c \right]\right\}, \quad (6)$$

where $i = 0, 1, \ldots N_p - 1$ is the main diagonal index, $\text{diag}\{\cdot\}$ refers to a diagonal matrix and $over = 2$ is the oversampling factor. Further on, we will refer to one diagonal element of this matrix product as $\beta(i)$. Now, we can rewrite (5) as,

$$\hat{\mathbf{h}} = \frac{\tilde{\mathbf{P}}^H \tilde{\mathbf{P}}}{\frac{\sigma^2}{N_c} \mathbf{C}_{\tilde{\mathbf{h}}}^{-1} + \tilde{\mathbf{P}}^H \tilde{\mathbf{P}}} \tilde{\mathbf{P}}^{-1} \hat{\mathbf{m}}_z, \quad (7)$$

we notice that everything preceding the inverse of the pilot matrix $\tilde{\mathbf{P}}$ can be presented by a diagonal matrix $\mathbf{X}_{diag}$, defined by its diagonal elements as

$$\mathbf{X}_{diag}[i,i] = \frac{\beta(i)}{\sigma^2/(N_c \sigma_{\tilde{h}(i)}^2) + \beta(i)}, \quad (8)$$

where $\sigma_{\tilde{h}(i)}^2$ is the power of the ith apriori channel tap estimate. The channel estimator can be stated as $\hat{\mathbf{h}} = \mathbf{X}_{diag} \tilde{\mathbf{P}}^{-1} \hat{\mathbf{m}}_z$. It is important to notice that the diagonal elements of $\mathbf{X}_{diag}$ are real and deterministic.

After obtaining the channel estimates a $4S$-point DFT of the channel estimate obtained $\hat{\mathbf{H}} = DFT\{\hat{\mathbf{h}}\}$, where S is the number of subbands in the synthesis bank, and in our simulations is set to be $S = 128$. The SCE is performed based on the frequency domain equivalent channel estimates, and a 3-tap complex FIR filter is used for SCE as in [4].

A. Theoretical MSE Limits for Ideal LMMSE Channel Estimator with OCI Training Sequences

Here we derive the analytic MSE threshold for the ideal LMMSE channel estimator, for which a block diagram is shown in Fig. 3. For ideal LMMSE channel estimator we assume that the LMMSE estimator has the true channel as apriori information. Also, in Fig. 3 the channel estimate and the user data symbols used for IC are the true ones. This setup is referred as estimation with IF.

The cyclic mean estimate vector of the received samples can be written in a matrix format as $\hat{\mathbf{m}}_z = \tilde{\mathbf{P}}\tilde{\mathbf{h}} + \mathbf{M}_d\tilde{\mathbf{h}} + \hat{\mathbf{m}}_w + \tilde{\mathbf{P}}\mathbf{h}_{aliasing}$, where $\hat{\mathbf{m}}_w$ is the cyclic mean estimate of the noise samples and $\mathbf{M}_d$ is a matrix of the cyclic mean estimates of the user data symbols and has a similar structure as $\mathbf{P}$.

Now, we can rewrite the channel estimator as

$$\hat{\mathbf{h}} = \mathbf{X}_{diag}[\tilde{\mathbf{h}} + \tilde{\mathbf{P}}^{-1}\mathbf{M}_d\tilde{\mathbf{h}} + \tilde{\mathbf{P}}^{-1}\hat{\mathbf{m}}_w + \mathbf{h}_{aliasing}]. \quad (9)$$

If we now define the estimation error, $\mathbf{e} = \hat{\mathbf{h}} - \tilde{\mathbf{h}}$, we can derive the MSE for the LMMSE estimator to be

$$\begin{aligned}
\sigma_e^2 &= E\{\|\mathbf{e}\|^2\} = trE\{\mathbf{e}\mathbf{e}^H\} = \cdots \\
&= trE\{(\mathbf{X}_{diag} - \mathbf{I})\tilde{\mathbf{h}}\tilde{\mathbf{h}}^H(\mathbf{X}_{diag} - \mathbf{I})^H\} \\
&\quad + trE\{\mathbf{X}_{diag}\tilde{\mathbf{P}}^{-1}\mathbf{M}_d\tilde{\mathbf{h}}\tilde{\mathbf{h}}^H\mathbf{M}_d^H\tilde{\mathbf{P}}^{-H}\mathbf{X}_{diag}^H\} \\
&\quad + trE\{\mathbf{X}_{diag}\tilde{\mathbf{P}}^{-1}\hat{\mathbf{m}}_w\hat{\mathbf{m}}_w^H\tilde{\mathbf{P}}^{-H}\mathbf{X}_{diag}^H\} \\
&\quad + trE\{\mathbf{X}_{diag}\mathbf{h}_{aliasing}\mathbf{h}_{aliasing}^H\mathbf{X}_{diag}^H\}.
\end{aligned} \quad (10)$$

All the cross terms inside the expectation are evaluated to zero, because of the normal independence assumptions between user data, channel response and additive noise. Because we have assumed discontinuous transmission, there is an additional weighting factor δ for the interference caused by the user data, as was shown in [5]. This weighting factor includes the knowledge of missing interfering user symbols in the beginning of the transmission. The weighting factor is defined as

$$\delta(m, i) = \begin{cases} \sigma_d^2/(overN_c) & \text{, if } m - i \geq 0 \\ \sigma_d^2(N_c - 1)/(overN_c^2) & \text{, otherwise,} \end{cases}, \quad (11)$$

where $m, i = 0, 1, 2, \ldots, overN_p - 1$ and $over = 2$ is the oversampling factor used in the receiver frontend. Based on the above derivation, the MSE of the ideal LMMSE estimator can now be written as
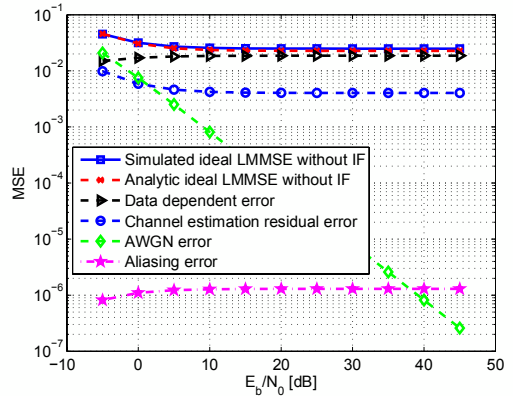

Fig. 4. Analytic and simulated MSE with 16-QAM constellation for ideal LMMSE channel estimator without IF.

$$\begin{aligned}
\sigma_{e, i-LMMSE}^2 &= \sum_{m=0}^{2N_p - 1} \frac{1}{(\frac{\sigma^2}{N_c}\sigma_{\tilde{h}(m)}^{-2} + \beta(m))^2} \\
&\quad \times \left[\frac{\sigma^4}{N_c^2\sigma_{\tilde{h}(m)}^2} + \left(\sum_{i=0}^{2N_p - 1} \sigma_{\tilde{h}(i)}^2\delta(m, i) \right. \right. \\
&\quad \left. \left. + \frac{\sigma_w^2}{N_c} + |h_{aliasing}(m)|^2\beta(m) \right) \beta(m) \right].
\end{aligned} \quad (12)$$

In the given MSE equation above, the summation indices go to $2Np - 1$ because of the 2 times oversampling. In (12), the first term of the four additive terms is referred as the residual channel estimation error, second term is related to the interference caused by the user data, third term is the error caused by the AWGN noise and the fourth term is the aliasing error caused by the usage of shorter channel estimator than the true equivalent channel.

For the simulated results presented in this paper, we have channel estimator of length 96 samples, while the true equivalent channel length is 142 samples, as in [5]. Therefore, the presented MSE results model the error between the channel estimate and the estimated portion of the true equivalent channel. The overall error between the whole equivalent channel and short channel estimate can also be easily approximated by adding the channel power outside the channel estimator to the obtained MSE, as discussed in [5].

The used frame length is 3840 symbols (7680 samples). Thus, we have 80 copies of OCI pilot sequence in top of each frame. For the presented MSE results, we assumed $\gamma = 0.26$ for the 16-QAM constellation. The AWGN noise power was assumed to be known in the receiver.

In Fig. 4 we have presented the simulated and analytic MSE performance of the presented ideal LMMSE channel estimator without IF. In Fig. 4, the different error sources are plotted with separate lines. The simulated MSE follows closely the analytic
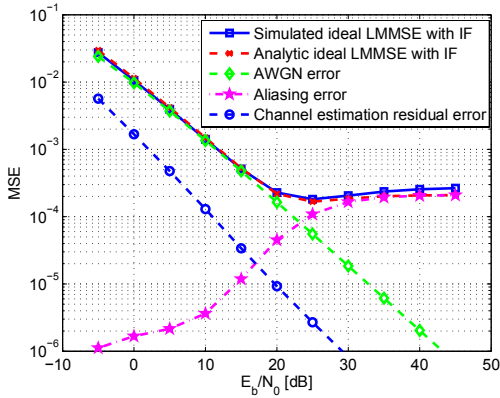
Fig. 5. Analytic and simulated MSE with 16-QAM constellation for ideal LMMSE channel estimator with IF.

one. We can see that the error performance is dominated by the interference caused by user data, as was to be expected. In all presented figures, x-axis follows logarithmic E_b/N_0 values, where E_b/N_0 is the bit energy to one sided noise power spectral density ratio.

In Fig. 5, the performance of the ideal LMMSE channel estimator with IF is presented. The IF is studied because it provides the lower bound for the MSE for any hard symbol feedback scheme. The MSE for the ideal LMMSE estimator with IF is given as

$$
\sigma_{e,i-LMMSE,IF}^2 = \sum_{m=0}^{2N_p-1} \frac{1}{(\frac{\sigma^2}{N_c}\sigma_{\tilde{h}(m)}^{-2} + \beta(m))^2} \left[\frac{\sigma^4}{N_c^2\sigma_{\tilde{h}(m)}^2} + \left(\frac{\sigma_w^2}{N_c} + |h_{aliasing}(m)|^2\beta(m) \right) \beta(m) \right]. \tag{13}
$$

Based on the definition, this result does not include the error caused by the user data symbols because they are ideally removed. Interesting in the results is that the aliasing error is greatly enhanced in the IF case at high E_b/N_0 values. This causes an clear error floor in the MSE, which is similar as with ML channel estimation (see [5] for comparison). The aliasing error approaches the squared norm of the vector modelling aliasing, $\|\mathbf{h}_{aliasing}\|^2$, because with IF at high E_b/N_0 values the multiplier of the term $|h_{aliasing}(m)|^2$ approaches one.

B. Theoretical Limits for Combined ML-LMMSE Channel Estimator with OCI Pilot Sequences

Now, we assume that before the LMMSE channel estimator we have an ML channel estimator [5], which is used to provide the apriori channel estimates used by the LMMSE estimator. The channel estimator structure is shown in Fig. 6 and is referred throughout this paper as combined ML-LMMSE channel estimator.

We utilise the equations (12) and (13) to obtain the MSE estimates for the combined structure. Here we assume that

the error and the channel estimate from the ML channel estimator are independent and that the average estimation error is evenly spread over all estimated channel taps. Thus, we can approximate the power of each estimated channel tap by

$$
E|\hat{h}_{ML}(i)|^2 = \sigma_{\tilde{h}(i)}^2 + \frac{\sigma_{ML,\mathbf{e}}^2}{overN_p}, \tag{14}
$$

where $\sigma_{ML,\mathbf{e}}^2$ is the MSE of the ML channel estimator. It is defined as [5]

$$
\sigma_{ML,e}^2 = \sum_{m=0}^{2N_p-1} \frac{1}{\beta(m)} \left[\sum_{i=0}^{2N_p-1} \delta(m,i)\sigma_{\tilde{h}(i)}^2 + \sigma_w^2/N_c \right] + \|\mathbf{h}_{aliasing}\|^2, \tag{15}
$$

without IF, and as [5]

$$
\sigma_{ML,e,IF}^2 = \frac{\sigma_w^2}{N_c} \sum_{m=0}^{2N_p-1} \frac{1}{\beta(m)} + \|\mathbf{h}_{aliasing}\|^2, \tag{16}
$$

with IF. These equations are used to derive the results shown in Figures 7 and 8.

Based on these assumptions, the equations (12) and (13) can be rewritten as

$$
\sigma_{e,ML-LMMSE}^2 = \sum_{m=0}^{2N_p-1} \frac{1}{(\frac{\sigma^2}{N_c}(\sigma_{\tilde{h}(m)}^2 + \frac{\sigma_{ML,e}^2}{overN_p})^{-1} + \beta(m))^2}
$$
$$
\times \left[\frac{\sigma^4}{N_c^2}(\sigma_{\tilde{h}(m)}^2 + \frac{\sigma_{ML,e}^2}{overN_p})^{-1} + \left(\sum_{i=0}^{2N_p-1} \sigma_{\tilde{h}(i)}^2\delta(m,i) + \frac{\sigma_w^2}{N_c} + |h_{aliasing}(m)|^2\beta(m) \right) \beta(m) \right], \tag{17}
$$

and

$$
\sigma_{e,ML-LMMSE,IF}^2 = \sum_{m=0}^{2N_p-1} \frac{1}{(\frac{\sigma^2}{N_c}(\sigma_{\tilde{h}(m)}^2 + \frac{\sigma_{ML,e,IF}^2}{overN_p})^{-1} + \beta(m))^2}
$$
$$
\times \left[\frac{\sigma^4}{N_c^2}(\sigma_{\tilde{h}(m)}^2 + \frac{\sigma_{ML,e,IF}^2}{N_p})^{-1} + \left(\frac{\sigma_w^2}{N_c} + |h_{aliasing}(m)|^2\beta(m) \right) \beta(m) \right]. \tag{18}
$$

The difference between the ideal LMMSE and ML-LMMSE with IF is that, although we fully remove the interference caused by user data with IF, the ML-LMMSE channel estimator uses the channel estimates obtained from ML channel estimator as apriori estimates and incorporates the estimation error related to ML channel estimation. Therefore, this provides us a lower bound for the MSE performance in a coded system with hard symbol feedback when using the ML-LMMSE channel estimator.

In Fig. 7, the MSE of the combined ML-LMMSE channel estimation structure without IF feedback is provided. It shows
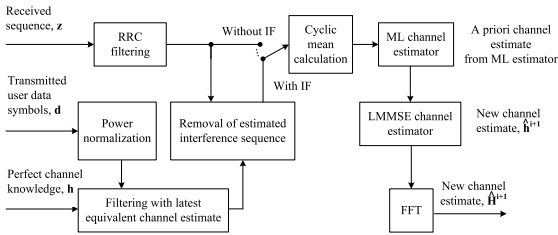
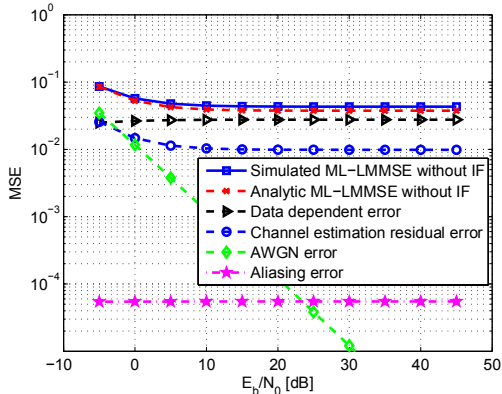Fig. 6. Combined ML-LMMSE channel estimator with IC based on IF.

Fig. 7. Analytic and simulated MSE with 16-QAM constellation for the combined ML-LMMSE channel estimator without IF.

that the used approximations provide good MSE estimation accuracy. The MSE is clearly higher than with ideal LMMSE, as was expected. Also, the channel estimation residual error is increased because of the ML estimation error. Even so, it does not dominate the MSE and is clearly lower than the interference caused by the user data symbols.
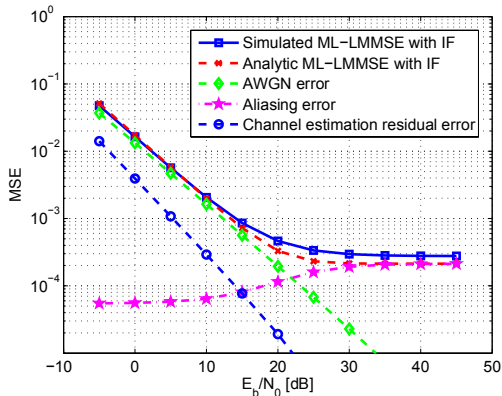

Fig. 8. Analytic and simulated MSE with 16-QAM constellation for the combined ML-LMMSE channel estimator with IF.

In Fig. 8, the performance of the combined channel estimator with IF is shown. Interestingly, we notice that with IF we can achieve almost the same performance as with ideal LMMSE. This indicates that through iterative IC processing we can achieve good performance.

IV. CONCLUSION

In this paper we have derived the MSE performance limits for the ideal LMMSE channel estimator and for the combined ML-LMMSE channel estimator, which is based on LMMSE using ML channel estimates as apriori information. The simulated MSE values follow well the presented analytical ones.

Based on the results, we can state that even though the performance of the combined channel estimation structure is degraded by the accuracy of the ML channel estimator, the ML-LMMSE solution can provide performance improvements through modest increment in the realisation complexity with higher order constellations. The channel estimation complexity is roughly doubled, but compared to the complexity of the whole reception chain, including e.g. some iterative channel decoding algorithm, this increase is minor.

The SI pilot structure studied in this paper provides interesting opportunities for future ad-hoc and device-to-device communications. In the future work, throughput performance comparison versus traditional time domain multiplexed pilots and analytical limits for the symbol error rate at the FB based maximum ratio combiner output are also of great interest.

REFERENCES

[1] S. Adireddy, L. Tong, and H. Viswanathan, "Optimal Placement of Known Symbols for Frequency-Selective Block-Fading Channels," *IEEE Trans. Inf. Theory*, vol. 48, pp. 2338–2353, August 2002.
[2] P. Hoeher and F. Tufvesson, "Channel Estimation with Superimposed Pilot Sequence," in *Global Telecommunications Conference 1999, GLOBECOM '99*, vol. 4, 1999, pp. 2162–2166.
[3] A. G. Orozco-Lugo, M. M. Lara, and D. C. McLernon, "Channel Estimation Using Implicit Training," *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 240–254, 2004.
[4] Y. Yang, T. Ihalainen, M. Rinne and M. Renfors, "Frequency-domain Equalization in Single-carrier Transmission: Filter Bank Approach," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, Article ID 10438, 16 pages, 2007.
[5] T. Levanen, J. Talvitie and M. Renfors, "Improved Performance Analysis for Super Imposed Pilot Based Short Channel Estimator," in *IEEE International Workshop on Signal Processing Advances for Wireless Communications, SPAWC 2010*, Marrakech, Morocco, 20-23 June 2010.
[6] T. B. Sorensen, P. E. Mogensen, and F. Frederiksen, "Extension of the ITU Channel Models for Wideband (OFDM) Systems," in *IEEE 62nd Vehicular Technology Conference 2005, VTC-2005-Fall*, Dallas, Texas USA, 25-28 September 2005, pp. 392–396.
[7] Y. Yang, T. Ihalainen and M. Renfors , "Filter Bank Based Frequency Domain Equalizer in Single Carrier Modulation," in *Proc. IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC '06*, September 2006.
[8] M. Pukkila, "Iterative Receivers and Multichannel Equalisation for Time Division Multiple Access Systems," Ph.D. dissertation, Helsinki University of Technology, Espoo, Finland, 10 October 2003, ISBN 951-22-6717-9.

Publication P3

T. Levanen, J. Talvitie, and M. Renfors, "Performance Evaluation of a DDST Based SIMO SC System with PAPR Reduction", in *The 6th International Symposium on Turbo Codes & Iterative Information Processing, ISTC 2010*, Brest, France, 6-10 Sept. 2010.

Performance Evaluation of a DDST Based SIMO SC System with PAPR Reduction

Toni Levanen, Jukka Talvitie and Markku Renfors
Department of Communications Engineering
Tampere University of Technology
P.O.Box 553, FIN-33101, Finland
Email: toni.levanen@tut.fi

Abstract—In this paper we concentrate on the symbol level peak to average power ratio (PAPR) increase caused by a data dependent superimposed pilot sequence. Because we add the pilot sequence on top of the user data symbols, the dynamic range of the transmitted signal may increase significantly. We propose the usage of a simple limiter in the transmitter and a hard symbol estimate based iterative estimator for the receiver. We show by simulations that if we allow a modest increase in the symbol level PAPR, the spectral efficiency of the data dependent superimposed pilot based system is better than the traditional time domain multiplexed pilot based system in a block fading channel.

Keywords: data dependent superimposed pilots; iterative receiver; peak to average power ratio; throughput comparison

I. INTRODUCTION

Channel estimation and equalisation are crucial parts of modern receiver architectures. As we aim for higher and higher spectral efficiencies, the number of time instances allocated for training in the traditional time domain multiplexed (TDM) systems should be minimised. At the moment, the superimposed (SI) pilots [1] are seen as a potential solution. SI pilots are added directly on top of the user data, and thus all time instances contain user data. In other words, by using SI pilots we can improve the spectral efficiency by allowing the user information to occupy the whole spectral region designed for communications. The downside is that the user information interferes greatly with the pilot sequence and that the user data symbol power to interference power ratio is decreased.

To overcome this problem of self interference, in [2] a data dependent superimposed training (DDST) method was presented. The basic idea is very simple. Because the cyclic pilot sequence has its energy concentrated on certain frequency bins, we set the user data frequency response to zero on these frequency bins. This equals with removing the cyclic mean of the user data symbol sequence in the time domain. Therefore there is no interference from the user data to the pilot symbols. This can be seen as frequency domain multiplexed (FDM) pilot based training, but the difference to the basic setup is that the signal spectrum is not widened because of the used SI training symbols. Similar approach was studied in [3], where

also the PAPR problem was discussed without any solutions to decrease the PAPR created by the SI pilots.

The DDST training method is suitable especially for wide band single carrier (SC) systems. With multicarrier systems this would mean that we lose some subcarriers for pilot symbols. The problem with the DDST is the increased peak to average power ratio (PAPR), which violates one of the main benefits of using SC transmission. We address this problem by simply limiting the peak amplitudes at symbol level before transmission. Then, in the receiver side, we have a simple feedback based on hard symbol estimates, which we use to estimate the missing cyclic mean and the limited amplitudes.

We have extended the model provided in [2] to our SC model with filter bank (FB) based receiver structure, presented in [4]. The channel estimates are obtained in time domain after which the sub-channel wise equalisation (SCE) is performed in the frequency domain (for more details, see [4] and references there in). The FB based receiver structure is used because it provides close to ideal linear equaliser performance, has good spectral containment properties and is equally applicable also to SC-FDMA (DFT-S-OFDMA) as used in 3GPP-LTE uplink.

This paper is structured as follows. First we present the used system model. Next, in Section III we briefly describe the used ML-LMMSE channel estimation scheme for DDST. In Section IV the throughput performance comparison of DDST and TDM training based systems is provided. Finally, in Section V conclusions and future topics are provided.

Notation: Superscripts T and H denote the transpose and Hermitian transpose operators, $\otimes$ refers to the Kronecker product and $\circ$ defines a continuous time convolution. For complex numbers $|z|$ defines the absolute value of z and $\arg(\cdot)$ gives the argument of a complex number. For a complex vector $\mathbf{z}$, $|\mathbf{z}| = [|z_0|, \ldots, |z_{N-1}|]$ defines an element wise absolute value operation. The statistical expectation is denoted by $E[\cdot]$. The $(N \times N)$ identity matrix is denoted by $\mathbf{I}_N$ and the $(M \times M)$ matrix of all ones by $\mathbf{1}_M$. Occasionally, $\mathbf{1}_M$ can refer also to a column vector of ones with length M, but this is always clear from the context or mentioned in the text. Finally, $\mathrm{diag}(\mathbf{a}) = \mathrm{diag}(a_0, \ldots, a_{N-1})$ is an $(N \times N)$ diagonal matrix whose nth main diagonal value is a_n. Matrices are denoted by boldface uppercase letters and vectors by boldface lowercase letters.

II. System Model

The system design originates from the uplink assumption. Thus, the complexity of the transmitting end is kept as small as possible and most of the complexity is positioned to the receiving end. The very simple block level design of the transmitter is given in Fig. 1. The transmitter contains bit source, channel encoder, interleaver (represented by π function), symbol mapper, pilot insertion, peak amplitude limiter and the transmitter pulse shape filter.

Let us assume that our symbol mapper produces a vector of data symbols $\mathbf{d}$ from some finite alphabet $\mathcal{A}^N$, where N is the frame (vector) length. We will use a pilot sequence, $\mathbf{p}$, which has length N_p. The pilot sequence is an optimal channel independent (OCI) sequence that was defined in [5]. In addition, we assume that our frame length is an integer multiple of N_p, given as $N = N_c N_p$, where N_c is the number of cyclic copies per frame. With the DDST, we first remove the cyclic mean of the data vector. As shown in [2], this can be represented as

$$\mathbf{z} = (\mathbf{I} - \mathbf{J}_{Tx})\mathbf{d}, \tag{1}$$

where $\mathbf{J}_{Tx} = (1/N_c)\mathbf{1}_{N_c} \otimes I_{N_p}$. Now the data dependent pilot sequence is given as $\mathbf{p}_d = -\mathbf{J}_{Tx}\mathbf{d}$. The symbol sequence including user data symbols, data dependent pilot sequence and the cyclic pilot sequence is given as $\mathbf{s} = \mathbf{d} + \mathbf{p}_d + \mathbf{p}_c = \mathbf{z} + \mathbf{p}_c$, where the cyclic pilot sequence is defined as $\mathbf{p}_c = \mathbf{1}_{N_c} \otimes \mathbf{p}$ and $\mathbf{1}_{N_c}$ is a vector of ones. This sequence, $\mathbf{s}$, is then inserted to the peak amplitude limiter from which the limited signal $\check{\mathbf{s}}$ is obtained. Finally, we normalise the signal to have unity power to obtain transmitted symbols $\tilde{\mathbf{s}}$. We define the power of the data sequence to be $\sigma_d^2 = 1 - \gamma$ and the power of the known pilot sequence to be $\sigma_{p_c}^2 = \gamma$, where γ is the power normalisation factor defining the allocated power for transmitted user data symbols and known pilot sequence.

The peak amplitude limiter takes as the maximum allowed amplitude value, a_{max}, the maximum amplitude value of the used constellation $\mathcal{A}$, defined as $\{a_{max} = \max(|(d)|), d \in \mathcal{A}, \sigma_d^2 = 1\}$. Note that here a_{max} is related to the maximum amplitude of a symbol constellation whose average power is normalised to unity. We chose this amplitude value to have similar peak powers as with time domain multiplexed (TDM) pilot based system. Now we can define the limited symbol sequence as

$$\check{s}(k) = \begin{cases} s(k), \text{ if } |s(k)| \leq a_{max}, \\ a_{max} \cdot exp(j\arg(s(k))), \text{ if } |s(k)| > a_{max}. \end{cases} \tag{2}$$

Now we have an amplitude limited symbol sequence whose PAPR is closer to the original value related to the data symbol sequence $\mathbf{d}$. The peak amplitudes are now limited to the original value, whereas the average power of the sequence is slightly decreased. In the simulations presented in Section IV, the energy per bit over one sided noise power spectral density, E_b/N_0, is defined for the normalised average power. In Table I the different average and peak powers with and without normalisation and the related PAPR are given for each constellation. Averaged powers σ_s^2 and $\sigma_{\tilde{s}}^2$ are obtained by averaging over 1000 frames.
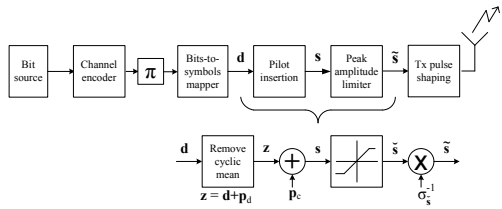

Fig. 1. Transmitter model.

TABLE I
SIMULATED POWER MEASURES FOR THE USED CONSTELLATIONS

	QPSK	16-QAM	64-QAM
σ_s^2	0.987	0.987	0.987
$\sigma_{\tilde{s}}^2$	0.795	0.946	0.980
$max(|s|)^2$	2.419	3.405	3.760
$max(|\check{s}|)^2$	1	1.8	2.33
$max(|\tilde{s}|)^2$	1.303	1.955	2.459
PAPR before limiter	2.451	3.464	3.805
PAPR after limiter	1.303	1.955	2.459
PAPR reduction (%)	46.8	43.6	35.4

By using a simple limiter in the transmitter we can remove most of the symbol level PAPR increase caused by cyclic mean removal and SI pilots. The QPSK modulation is the most sensitive to SI pilots because it originally has PAPR equal to one. Even after the limiter the symbol level PAPR is increased by 30.3%. 16-QAM and 64-QAM modulations are less sensitive and the remaining PAPR increase after the limiter is only 8.6% and 5.4%, respectively. Note that in this paper we have studied only the symbol level PAPR to provide some preliminary results on the limiter performance, and the true impact on the transmitted signal PAPR is left for future studies.

We define a vector $\mathbf{e}_{limiter} = \check{\mathbf{s}} - \mathbf{s}$, which contains the information removed by the limiter from the sequence $\mathbf{s}$. In other words, it represents an additive error sequence generated by the limiter. In this paper we are not concerned with the estimation of the variance of this noise based on other system parameters, but this is an interesting problem for future studies. The error caused by the limiter is simply assumed to be zero mean complex Gaussian noise for complex constellations.

We assume a discontinuous block wise transmission where the channel is assumed to be time invariant during the transmission time of one block. The used channel model is ITU-R Vehicular A channel with about 2.5 μs delay spread and approximately 20 MHz bandwidth [6].

In Fig. 2 we have presented a block diagram of our multiantenna receiver. We assume perfect synchronization in frequency and time domain, ideal down conversion and 2 times oversampling of the received signal in Rx block. Based on these ideality assumptions, we can present the channel between transmitter and receiver as a 2 times oversampled discrete time equivalent channel as $h(k) = |h_{Tx}(t) \circ h_{channel}(t) \circ$
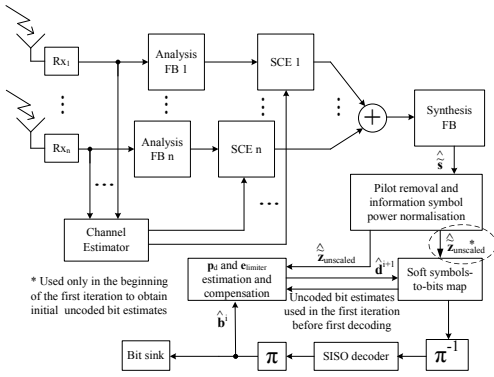
Fig. 2. Receiver model using multiantenna reception with maximum ratio combining and hard decision feedback loop for cyclic mean and limiter error estimation with DDST based channel estimation.

$h_{Rx}(t)|_{t=kT/2}$. The received symbol $y(k)$ can be given as

$$y(k) = \sum_{m=0}^{M-1} h(m)s(k-m) + w(k), \qquad (3)$$

where M is the channel length in samples, k is the time index for 2 times oversampled symbol sequence and $s(k)$ is a transmitted symbol, which is zero if $k < 0$ or $k > 2N-1$. The noise term $w(k) = |h_{Rx}(t) \circ v(t)|_{t=kT/2}$, where $v(t)$ is complex additive white Gaussian noise (AWGN), is simply modelled as AWGN without considering the correlation caused by the 2 times oversampled receiver pulse shape filtering. Because of the oversampling, $s(k) = d(k) = p_c(k) = 0$ when k modulus $2 = 1$. We will be more concentrated on the matrix notation of the signal model, which is given as

$$\mathbf{y} = \mathbf{H}\tilde{\mathbf{s}} + \mathbf{w} = \tilde{\mathbf{S}}\mathbf{h} + \mathbf{w}, \qquad (4)$$

where the matrix $\tilde{\mathbf{S}} = \mathbf{D} + \mathbf{P}_d + \mathbf{P}_c + \mathbf{E}_{limiter}$ is built from the user data symbols, data dependent pilot sequence, known cyclic pilot sequence and the additive error generated by the limiter, respectively.

Because we assume a discontinuous block wise transmission, all matrices $\mathbf{D}, \mathbf{P}_d, \mathbf{P}_c$ and $\mathbf{E}_{limiter}$ have the form

$$\begin{bmatrix} x_0 & 0 & \dots & 0 & 0 \\ x_1 & x_0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{N_p-1} & x_{N_p-2} & \dots & x_1 & x_0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{N-1} & x_{N-2} & \dots & x_{N-N_p+1} & x_{N-N_p} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & x_{N-1} \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix}, \quad (5)$$

including the zeros before and after the transmitted frame. Note that the matrices are now of dimension $(N + N_p \times N_p)$ and that we have assumed that $M = N_p$. This means that in

the receiver we have to do the cyclic mean calculation over $N_c + 1$ copies. Thus, the cyclic mean of the received sequence is given as

$$\begin{aligned} \hat{\mathbf{m}}_y &= \mathbf{J}_{Rx}\mathbf{y} \\ &= (\mathbf{J}_{Rx}\mathbf{D} + \mathbf{J}_{Rx}\mathbf{P}_d + \mathbf{J}_{Rx}\mathbf{P}_c + \mathbf{J}_{Rx}\mathbf{E}_{limiter})\mathbf{h} + \mathbf{J}_{Rx}\mathbf{w} \\ &= \mathbf{P}\mathbf{h} + \hat{\mathbf{M}}_{e_{limiter}}\mathbf{h} + \hat{\mathbf{m}}_w, \end{aligned}$$
$$(6)$$

where $\mathbf{J}_{Rx} = (1/N_c)\mathbf{1}_{N_c \times N_c+1} \otimes I_{N_p}$ and $\hat{\mathbf{m}}_x$ defines the approximated cyclic mean vector of vector $\mathbf{x}$. Here $\hat{\mathbf{M}}_{e_{limiter}}$ is a matrix built from different cyclic shifts of $\hat{\mathbf{m}}_{e_{limiter}}$ and $\mathbf{P}$ is a cyclic matrix built from the OCI pilot sequence $\mathbf{p}$.

From the receiver frontend, the oversampled signal is provided for channel estimator and for analysis FB. The channel estimation algorithm to be defined in Section III is for one receiver branch and is simply repeated for each diversity branch. Here the channel equalisation of different branches can be done by either parallel or sequential processing. After obtaining channel estimate, SCE is performed in the frequency domain. It should be noted that the equalisation is now performed within mildly frequency selective subbands. More details on the equaliser structure can be found from [4] and references there in.

After SCE, different antenna branches are added together subsignal wise according to the maximum ratio combining principle. The composite subsignals are then recombined in the synthesis FB, which also efficiently realises the 2 times down conversion of the sampling rate. After the synthesis FB, the received sequence power is normalised to $1 + \sigma_w^2$, which corresponds to the total received power. We have assumed that we exactly know the noise variance in the receiver. After received power normalisation we remove the cyclic mean of the received sequence and normalise it based on the pilot power allocation. Thus, we obtain an estimate for the $\mathbf{z}$ with cyclic mean equal to zero and including the limiter error, given as

$$\hat{\mathbf{z}} = \sqrt{\frac{1}{1-\gamma}}(\mathbf{I} - \mathbf{J}_{\mathbf{Tx}})\sqrt{\frac{1+\sigma_w^2}{\sigma_{\hat{s}}^2}}\hat{\mathbf{s}}. \qquad (7)$$

We have used the same notation for the vectors containing the same data but with different sampling rate. This is for the sake of clarity and should not cause any problems for the reader. The vectors before the analysis filter banks in Fig. 2 are related to 2 times oversampled sequences and the vectors after the synthesis filter bank are related to sequences sampled at the symbol frequency.

Next, we generate initial hard symbol estimates based on the *Soft symbols-to-bits* block output and use them for initial $\mathbf{p}_d$ and $\mathbf{e}_{limiter}$ estimation in the $\mathbf{p}_d$ *and* $\mathbf{e}_{limiter}$ *estimation and compensation* block. This block is presented in more detail in Fig. 3. This uncoded estimation of the cyclic mean and data dependent pilot signal is similar to one iteration of the hard, uncoded feedback algorithm proposed already in [2]. As a result, we have rough estimates of $\sigma_{\hat{s}}$, $\mathbf{p}_d$ and $\mathbf{e}_{limiter}$ available in the second soft symbols-to-bits mapping. Note that everything described above takes place before the first soft decoding process. This preprocessing phase causes
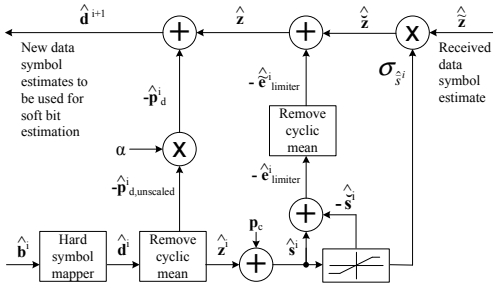
Fig. 3. A block diagram presenting the operations performed inside the $\mathbf{p}_d$ and $\mathbf{e}_{limiter}$ *estimation and compensation* block.

an insignificant increase in the receiver complexity, but does improve the quality of the first soft bit estimates provided for the soft decoder.

The soft bit estimates are then provided to the Turbo decoder (denoted as SISO decoder in Fig. 2). After the turbo decoding first bit estimates are provided for BER evaluation and this is considered to be the first iteration. The first iteration is not considered as a feedback iteration in our results.

In Fig. 3 subscript i refers to the iteration number and we have used notation $\hat{\tilde{\mathbf{z}}}$ to represent our estimates of the data symbol sequence with cyclic mean set to zero and including the limiter error obtained from the *Pilot removal and information symbol power normalisation* block. First we generate hard symbol estimates based on the latest bit estimates $\hat{\mathbf{b}}^i$. Then we calculate the symbol wise cyclic mean and remove it from the symbol sequence. Next, we add the known pilot sequence on top of the symbol sequence $\hat{\mathbf{z}}^i$ and provide this sequence to the amplitude limiter. Then we calculate the limiter error estimate based on the input and the output of the limiter function.

Because we used power scaling in the transmitter, we approximate this normalisation factor in the receiver and use it to obtain normalised estimates $\hat{\tilde{\mathbf{z}}} = \sigma_{\hat{\tilde{s}}^i}\hat{\tilde{\mathbf{z}}}$. Next, we remove the latest limiter error estimates with cyclic mean removed to get $\hat{\mathbf{z}} = \hat{\tilde{\mathbf{z}}} - \hat{\tilde{\mathbf{e}}}^i_{limiter}$. We remove the cyclic mean from the limiter error estimates because it was also removed from our symbol estimates $\hat{\tilde{\mathbf{z}}}$ in the *Pilot removal and information symbol power normalisation* block.

Next, we remove an normalised version of the latest data dependent pilot sequence $\hat{\mathbf{p}}^i_d = \alpha \hat{\mathbf{p}}^i_{d,unscaled}$ from $\hat{\mathbf{z}}$ to finally obtain our new unscaled data symbol estimates $\hat{\mathbf{d}}^{i+1}$. The power scaling factor for data dependent pilot sequence is given as

$$\alpha = \sqrt{\frac{E[|p_d|^2]}{\sigma^2_{\hat{p}^i_{d,unscaled}}}} = \sqrt{\frac{\sigma^2_d/N_c}{\sigma^2_{\hat{p}^i_{d,unscaled}}}}, \qquad (8)$$

and it is used because typically during the first iterations the estimated cyclic mean has a too small variance. We know that the expected variance of the cyclic mean should be $E[|p_d|^2] = \sigma^2_d/N_c$, and therefore we use this normalisation factor to improve the performance.

III. ML-LMMSE CHANNEL ESTIMATION

When defining the LMMSE channel estimator, we want to minimise the expected value of the squared error, $E\{|\mathbf{h} - \hat{\mathbf{h}}|^2\}$. If we now make the assumptions that the noise and the total interference experienced by the pilot sequence is AWGN, channel taps are i.i.d. and have zero mean, i.e. $E\{\mathbf{h}\} = \mathbf{0}$, the LMMSE estimator can be simplified to [7]

$$\hat{\mathbf{h}} = (\sigma^2_w \mathbf{C}^{-1}_{\mathbf{h}_{apriori}} + \mathbf{P}^H_c \mathbf{P}_c)^{-1} \mathbf{P}^H_c \mathbf{y}, \qquad (9)$$

where σ^2_w is the AWGN channel noise. The channel covariance matrix $\mathbf{C}_{\mathbf{h}_{apriori}}$, contains the apriori information of the channel tap values. The apriori information of the channel taps are obtained through a ML channel estimator, defined as

$$\hat{\mathbf{h}}_{ML} = \frac{\mathbf{P}^H}{N_p \sigma^2_p} \hat{\mathbf{m}}_y. \qquad (10)$$

By the assumption of independent tap coefficients, it becomes diagonal, i.e., $\mathbf{C}_{\hat{\mathbf{h}}_{ML}} = diag\{|\hat{h}(0)|^2, |\hat{h}(1)|^2, \cdots, |\hat{h}(N_p - 1)|^2\}$. By assuming the cyclic OCI training sequence, the LMMSE estimator can be reduced to

$$\hat{\mathbf{h}}_{ML-LMMSE} = \left(\frac{\sigma^2_w}{N_c} diag(|\hat{\mathbf{h}}_{ML}|^{-2}) + N_p \sigma^2_p I_{N_p}\right)^{-1} \mathbf{P}^H \hat{\mathbf{m}}_y. \qquad (11)$$

With DDST, the ML channel estimate is quite good, but we can further improve it by using the LMMSE channel estimator following the ML channel estimator. For this reason, the channel estimator is named as ML-LMMSE. Another reason to use ML-LMMSE structure is to compensate for the additional error caused by the limiter error present in the signal. It would probably improve the estimation performance if we could provide an limiter error variance estimate for the channel estimator, but this is left for future studies. This kind of channel estimator structure with traditional SI pilots and iterative interference cancelling feedback was studied in [8].

IV. SIMULATED THROUGHPUT PERFORMANCE COMPARISON BETWEEN DDST AND TDM BASED TRAINING

The used channel model is ITU-R Vehicular A channel with about 2.5 μs delay spread and approximately 20 MHz bandwidth [6]. The delay spread is 78 samples in the receiver, where 2 times oversampling is used in the analysis filter bank. The oversampling allows us to efficiently realise the RRC filtering in frequency domain combined in the channel equalisation process. More details can be found, e.g., [4] and references there in.

The channel codec is a turbo codec with generator matrix $G = \begin{bmatrix} 1 & \frac{15}{13} \end{bmatrix}$. The used interleavers are bitwise S-interleavers, where the distance parameter is defined as $S = \sqrt{L/2}$, where L is the length of the unit which is interleaved. In channel interleaving the unit is the whole transmitted frame and inside a turbo encoder/decoder the interleaving unit one code block. Each transmitted frame is divided into Q coded blocks, where 2^Q defines the used constellation size.

Symbol rate	15.36 MHz
Signal bandwidth	18.74 MHz
Frame duration	250 μs
Order of the RRC filter	32
RRC roll-off	0.22
Symbols per frame	3840
TDM pilot symbols per frame	384
γ with 4-QAM	0.1
γ with 16-QAM	0.05
γ with 64-QAM	0.02
No. of subbands in the analysis FB	256
No. of subbands in the synthesis FB	128
Overlapping factor	5
FB roll-off	1

We have run the simulations for QPSK, 16-QAM and 64-QAM constellations with code rates $R = 0.5$, $R = 0.67$ and $R = 0.75$. Puncturing is performed over parity bits. Some additional simulation parameters related to the simulation model are given in Table II. For DDST, the SI pilot powers were defined for each constellation by choosing the pilot power leading to the smallest average BER with three feedback iterations (in total four iterations). The averaging was done over all code rates and with 2 and 4 receiving antennas. These results could have been further optimized by defining different pilot powers for each coding and receiving antenna number pair, but this is out of the scope of this paper.

In the Fig. 4 we have presented results for DDST with combined ML-LMMSE after three feedback iterations with increased symbol level PAPR and for TDM using also ML-LMMSE type equaliser. Furthermore, in the presented spectral efficiency figures, we have presented the spectral efficiency with our system if the channel response is known in the receiver and no pilots are transmitted. This represents an upper bound of the spectral efficiency for the given system. The provided results assume that the receiver knows exactly the noise variance and are derived by choosing the highest achievable rate among all code-constellation pairs for each E_b/N_0 point.

From Fig. 4 we can clearly see how the DDST based system improves the spectral efficiency if we allow increased symbol level PAPR. It should be noted, that the symbol level PAPR does not directly map to PAPR after the pulse shape filtering and this topic requires additional studies. The maximum spectral efficiency difference for each constellation is equal to 10%, which corresponds to the number of pilot symbols allocated for TDM. We could have used less TDM pilots with 4 receiving antennas, but we thought that this would not correspond to a real life scenario, where the transmitted frame structure is fixed regardless of the number of receiving or transmitting antennas.

V. CONCLUSION

In this paper we have presented a simple limiter approach to decrease the severe PAPR problem related to DDST in the transmitter. For the reception, a simple algorithm to estimate
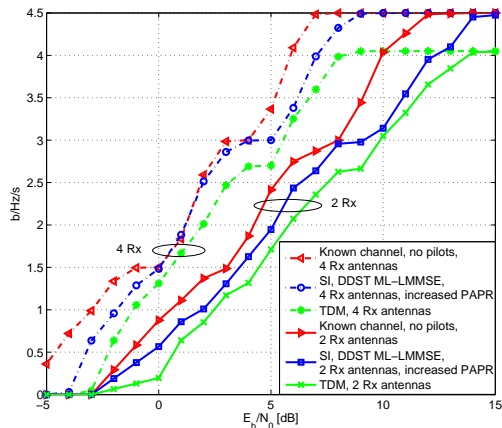

Fig. 4. Throughput performance comparison for DDST and TDM training based systems. In addition, performance without any pilots and known channel response is shown.

the additional error term caused by the limiter was presented. With the given system, spectral efficiency improvements can be achieved if we allow increased symbol level PAPR in the transmitter and increased reception complexity in the receiver. These results are preliminary, but they provide insight to the possibilities of DDST training in wireless communications.

In a coming article, the MSE performance bounds of the ML-LMMSE estimator with DDST will be presented. Further more, additional details on the the true transmitted signal PAPR, limiter error structure and their effects on the system performance will be provided.

REFERENCES

[1] P. Hoeher and F. Tufvesson, "Channel Estimation with Superimposed Pilot Sequence," in *Global Telecommunications Conference 1999, GLOBECOM '99*, vol. 4, 1999, pp. 2162–2166.

[2] M. Ghogho et al., "Channel Estimation and Symbol Detection for Block Transmission Using Data-Dependent Superimposed Training," *IEEE Signal Process. Lett.*, vol. 12, no. 3, pp. 226–229, 2005.

[3] C. Lam et al., "Channel Estimation for SC-FDE Systems Using Frequency Domain Multiplexed Pilots," in *IEEE 64th Vehicular Technology Conf., 2006, VTC-2006 Fall*, Montral, Canada, 25-28 Sept., pp. 1–5.

[4] Y. Yang, T. Ihalainen, M. Rinne and M. Renfors, "Frequency-domain Equalization in Single-carrier Transmission: Filter Bank Approach," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, Article ID 10438, 16 pages, 2007.

[5] A. G. Orozco-Lugo, M. M. Lara, and D. C. McLernon, "Channel Estimation Using Implicit Training," *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 240–254, 2004.

[6] T. B. Sorensen, P. E. Mogensen, and F. Frederiksen, "Extension of the ITU Channel Models for Wideband (OFDM) Systems," in *IEEE 62nd Vehicular Technology Conference 2005, VTC-2005-Fall*, Dallas, Texas USA, 25-28 Sept. 2005, pp. 392–396.

[7] M. Pukkila, "Iterative Receivers and Multichannel Equalisation for Time Division Multiple Access Systems," Ph.D. dissertation, Helsinki University of Technology, Espoo, Finland, 10 Oct. 2003, ISBN 951-22-6717-9.

[8] T. Levanen and M. Renfors, "Improved Performance Bounds for Iterative IC LMMSE Channel Estimator with SI Pilots," in *21st Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Istanbul, Turkey, 26-30 Sept. 2010.

Publication P4

T. Levanen, J. Talvitie, and M. Renfors, "Throughput Evaluation of Time-Multiplexed and Data-Dependent Superimposed Training Based Transmission with Practical Power Amplifier Model", in *Eurasip Journal on Wireless Communications and Networking* special issue on *Algorithm and Implementation Aspects of Channel Codes and Iterative Receivers*, Vol. 49, 2012, doi:10.1186/1687-1499-2012-49.

RESEARCH   Open Access

Performance evaluation of time-multiplexed and data-dependent superimposed training based transmission with practical power amplifier model

Toni Levanen[*], Jukka Talvitie and Markku Renfors

Abstract

The increase in the peak-to-average power ratio (PAPR) is a well known but not sufficiently addressed problem with data-dependent superimposed training (DDST) based approaches for channel estimation and synchronization in digital communication links. In this article, we concentrate on the PAPR analysis with DDST and on the spectral regrowth with a nonlinear amplifier. In addition, a novel Gaussian distribution model based on the multinomial distribution for the cyclic mean component is presented. We propose the use of a symbol level amplitude limiter in the transmitter together with a modified channel estimator and iterative data bit estimator in the receiver. We show that this setup efficiently reduces the regrowth with the DDST. In the end, spectral efficiency comparison between time domain multiplexed training and DDST with or without symbol level limiter is provided. The results indicate improved performance for DDST based approaches with relaxed transmitter power amplifier requirements.

Keywords: channel estimation, data-dependent superimposed pilots, iterative receiver, nonlinear power amplifier, peak-to-average power ratio, spectral efficiency.

1 Introduction

Channel estimation and equalization are crucial parts of modern digital transmission links. As we aim for higher spectral efficiencies, the number of time instances allocated for training in the traditional time-domain multiplexed training (TDMT) systems should be minimized. At the moment, the superimposed (SI) scheme is a serious candidate for circumventing this issue, see for example [1-3] and references therein. SI pilots are added directly on top of the user data, and thus all time instances over the whole allocated spectral region contain user information. The downside is that the user information interferes greatly with the pilot sequence, increasing the mean squared error (MSE) of the initial channel estimates. Furthermore, the peak-to-average power ratio (PAPR) is considerably increased and the user-data-symbol-to-interference power ratio is decreased in detection.

To overcome this problem of self-interference (interference from the user data symbols in channel estimation), a data-dependent superimposed training (DDST)

scheme was presented in [4,5]. The basic idea is very simple. Because the cyclic pilot sequence has its energy concentrated on certain frequency bins, we set the user data frequency response to zero on these frequency bins. This is equivalent to removing the cyclic mean of the user data symbol sequence in the time domain. Therefore, there is no interference from the user data to the pilot symbols. Because the interference from the user data symbols is removed, DDST requires clearly lower pilot powers than traditional SI training to obtain the desired channel estimation MSE levels. This can also be seen as frequency-domain multiplexed (FDM) pilot based training, but the difference to the traditional approach is that the signal spectrum is not widened because of the used SI training symbols. With multicarrier systems, spectral nulling means that we lose some subcarriers for pilot symbols. Recently, a solution to circumvent this problem in multicarrier communications by the so called symbol blanking method was proposed in [6].

The DDST is suitable especially for wide-band single-carrier (SC) systems. The problem to be addressed in this article regarding the addition of DDST sequences is the increased peak power (PP) and PAPR, which violates one

* Correspondence: toni.levanen@tut.fi
Department of Communications Engineering, Tampere University of Technology, P.O. Box 553, FIN-33101, Finland

Springer

of the main benefits of using SC transmission. With increased PAPR we can expect increased spectral regrowth with nonlinear amplifiers, which are preferred in the mobile devices because of their higher efficiency. Based on the authors best knowledge, the effects of increased PP or PAPR on the spectral regrowth have not been taken into account in the recent literature in the performance comparisons between DDST and TDMT systems. More traditional SI-based training was studied in [7], where the frequency bins were in some cases nulled for improved channel estimation performance. The PAPR problem was discussed without any solutions to decrease the PAPR created by the SI pilots. We will address this problem by simply limiting the peak amplitudes at the symbol level before transmission. From now on, this symbol level amplitude limited DDST is denoted as LDDST.

In the receiver side, we have a simple feedback loop based on soft symbol estimates, which we use to estimate the missing cyclic mean and the limited amplitudes. In [8], we studied the symbol level PAPR and used an iterative receiver structure without any knowledge of the error generated by the symbol level amplitude limiter in the transmitter. In this article we will utilize the scaling information available based on Gaussian modeling of the data-dependent pilot sequence (cyclic mean) in the channel estimator.

This article is structured as follows. First we present the system model in Section 2. Then, in Section 3 we model the error caused by the symbol level limiter in the transmitted signal. Next, in Section 4 we briefly discuss the modifications used in the channel estimation algorithms because of the symbol level limiter. In Section 5, we concentrate on the symbol level PP and PAPR, on the PP and PAPR after the transmit pulse shape filtering, and show that the symbol level limiter can remove the PP increase and effectively reduce the PAPR. In addition, we discuss the spectral re-growth related to different training methods. In the Section 6, we provide improved iterative receiver algorithms taking into consideration the amplitude limiter in the transmitter and the removal of the data dependent pilots. Next, in Section 7, the throughput performance comparison of DDST and TDMT training based systems is provided. Finally, in Section 8, conclusions are provided.

Notation: Superscripts T and H denote the transpose and Hermitian transpose operators, $\otimes$ refers to the Kronecker product and o defines a continuous-time convolution. For complex numbers $|z|$ defines the absolute value of z and $\angle\cdot$ gives the argument of a complex number. In addition, Re(z) takes the real value of a complex number and Im(z) takes the imaginary value. Exponential function is noted by exp($\cdot$) and $\|z\|$ defines the Euclidean vector norm. The trace and statistical expectations are denoted by $tr[\cdot]$ and $E[\cdot]$. Rounding to the largest integer

not greater than x is given by the floor function $\lfloor x \rfloor$. The $(N \times N)$ identity matrix is denoted by $\mathbf{I}_N$ and the $(N \times M)$ matrix of all ones by $1_{N \times M}$. For oversampling, we define a column vector $\mathbf{r}$ with first element equal to one and i - 1 zeros after the first element, e.g., $\mathbf{r} = [1,0,...,0]^T$. We denote the length of this vector with r, which will represent the oversampling rate used in the receiver. Matrices are denoted by boldface uppercase letters and vectors by boldface lowercase letters. Finally, diag($\mathbf{a}$) = diag($a_1,...,a_n$) is an $(N \times N)$ diagonal matrix whose nth entry is a_n and diag($\mathbf{A}$) is a$(N \times 1)$ vector with values from the main diagonal of $\mathbf{A}$, which isa$(N \times N)$ square matrix.

2 System model

Our system design originates from the uplink assumption. Thus, the complexity of the transmitting end is kept as small as possible and most of the complexity is positioned to the receiving end. The block level design of the transmitter is given in Figure 1. The transmitter contains a bit source, channel encoder, interleaver (represented by π function), symbol mapper, pilot insertion, symbol level amplitude limiter, $L(\cdot)$, the transmitter pulse shape filter and nonlinear amplifier, $G(\cdot)$.
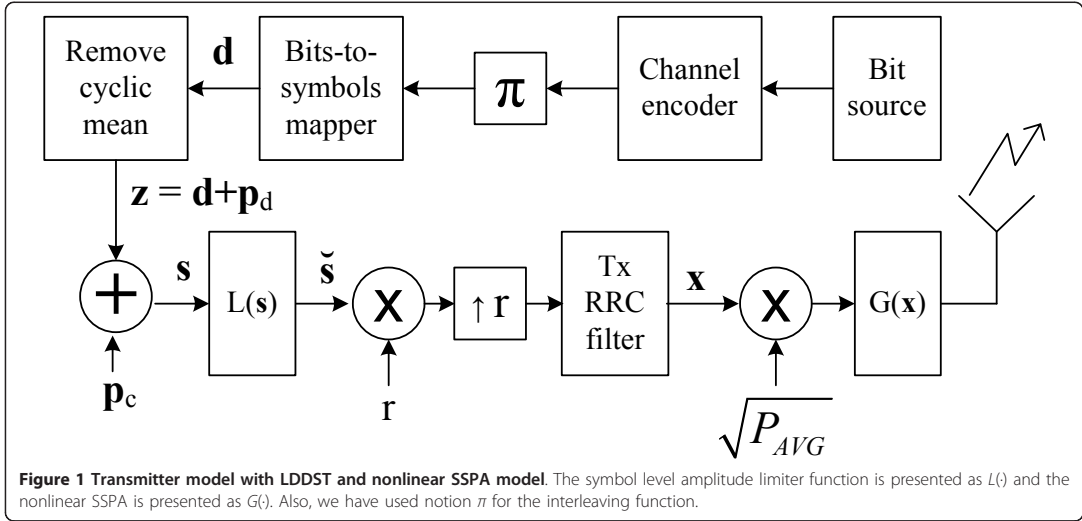
Let us assume that our symbol mapper produces a vector of data symbols d from some finite alphabet $\mathcal{A}^N$, where N is the frame (vector) length. We will use a pilot sequence, $\mathbf{p}$, which has length N_p. The pilot sequence is an optimal channel independent (OCI) sequence that was defined in [2], and rewritten here as

$$p(k) = \sigma_p e^{j\frac{\pi}{N_p}[k(k+v)]}, \tag{1}$$

where $k = 0,...,N_p$ - 1, $v = 1$ if N_p is odd and $v = 2$ if N_p is even number. In addition, we assume that our frame length is an integer multiple of N_p, given as $N = N_cN_p$, where N_c is the number of cyclic copies per frame. With the DDST, we first remove the cyclic mean of the data vector. As shown in [4], this can be expressed as

$$\mathbf{z} = (\mathbf{I} - \mathbf{J}_{Tx})\mathbf{d}, \tag{2}$$

where $\mathbf{J}_{Tx} = (1/N_c)1_{N_c \times N_c} \otimes \mathbf{I}_{N_p}$. Now the data dependent pilot sequence is given as $\mathbf{p}_d = -\mathbf{J}_{Tx}\mathbf{d}$. The data dependent pilot sequence is added on top of the data sequence in order to remove the cyclic mean of the data sequence, thus removing the interference caused by data sequence on the known pilot sequence. The symbol sequence including user data symbols, data dependent pilot sequence and the cyclic pilot sequence is given as $\mathbf{s} = \mathbf{d} + \mathbf{p}_d + \mathbf{p}_c = \mathbf{z} + \mathbf{p}_c$, where the cyclic pilot sequence is defined as $\mathbf{p}_c = 1_{N_c \times 1} \otimes \mathbf{p}$. For a more detailed explanation on DDST, see for example [9] and references therein. The symbol sequence, $\mathbf{s}$, is then inserted to the peak amplitude

Figure 1 Transmitter model with LDDST and nonlinear SSPA model. The symbol level amplitude limiter function is presented as $L(\cdot)$ and the nonlinear SSPA is presented as $G(\cdot)$. Also, we have used notion π for the interleaving function.

limiter from which the limited signal $\breve{s}$ is then obtained. This sequence is then oversampled with rate r, given as $\breve{s}_r = r\breve{s} \otimes \mathbf{r}$, and inserted to the transmit pulse shape filter to obtain transmitted sequence $\mathbf{x}$. We define the power of the data sequence to be $\sigma_d^2 = 1 - \gamma$ and the power of the known pilot sequence to be $\sigma_{pc}^2 = \gamma$, where γ is the pilot power allocation factor.

The peak amplitude limiter is presented by a function $L(\cdot)$, which takes as the maximum allowed amplitude value, $a_{\max}$, the maximum amplitude value of the used constellation $\mathcal{A}$, defined as $\{a_{\max} = \max(|(d)|), d \in \mathcal{A}, \sigma_d^2 = 1\}$. We use this value because we wanted to achieve similar type of PAPR behavior as with TDMT and that the limiter affects mainly pilot sequences added on top of the user data. The limited symbol sequence can be defined as

$$\breve{s}(k) = L(s(k)) = \begin{cases} s(k), & \text{if } |s(k)| \leq a_{\max}, \\ a_{\max} \cdot \exp(j\angle s(k)), & \text{if } |s(k)| > a_{\max}. \end{cases} \quad (3)$$

Now we have an amplitude limited symbol sequence whose PP is limited to the same value as the original data symbol sequence $\mathbf{d}$. The average power decrease, and the remaining PAPR increase, depends on the constellation. This kind of amplitude limiter, which keeps the argument difference between input and output as a constant, realizes so-called amplitude-modulation to amplitude-modulation (AM-AM) conversion [10], meaning that $|L(s(k))|$ depends only on $|s(k)|$.

We have chosen to study the hard limiting of the transmitted symbols, but of course other limiters with different input-output mappings require more studies. Furthermore, we have chosen to study symbol level

limiting instead of limiting the output of the Tx pulse shape filter, which is a more common approach for controlling the PAPR in SC transmission. From the literature concerning studies on PAPR with OFDM modulation, one can find several possible topics of study in order to reduce PAPR in DDST with a modified data-dependent pilot sequence, and these are left for future studies.

Let us define an error vector $\mathbf{e}_{\text{limiter}} = \breve{\mathbf{s}} - \mathbf{s}$, which contains the information removed by the limiter from the sequence $\mathbf{s}$. It represents an additive error sequence generated by the limiter. This model is used when we present the receiver feedback structure in Section 7.

The signal after the symbol level limiter, $\breve{\mathbf{s}}$, is then fed to the transmit pulse shape filter after over-sampling. We have used traditional root-raised-cosine (RRC) filtering with rolloff factor $\rho = 0.1$ and filter order $N_{\text{RRC}} = 64$. We have chosen two different scenarios for simulations. For the PAPR and spectral leakage simulations we have used four times oversampling, $r = 4$, and for the performance evaluations we have used two times oversampling, $r = 2$. We have chosen this setup for better understanding of the spectral spreading and because the used filter bank (FB) based equalizer is designed to work with two times oversampled sequences.

The nonlinear power amplifier model is a widely-used basic model, based on solid-state power amplifier (SSPA) model by Rapp [11]. The AM-to-AM conversion function for an input amplitude A is given as

$$G(A) = v \frac{A}{\left(1 + \left[\frac{vA}{A_0}\right]^{2p}\right)^{-2p}}, \quad (4)$$

where v is the small signal amplification, A_0 is the saturation amplitude of the amplifier and p defines the smoothness of the transition from linear region to the limiter region. The actual values chosen for the simulations are discussed in more detail in Section 7.

Based on Bussgang's theorem [12], we model the output of the power amplifier as $G(\mathbf{x}) = \alpha\sqrt{P_{\text{AVG}}}\mathbf{x} + \mathbf{n}_G$, where α is a scaling factor for the input signal, P_{AVG} is the average power of the transmitted frame, and $\mathbf{n}_G$ is uncorrelated Gaussian noise vector caused by the nonlinear power amplifier $G(\cdot)$. P_{Avg} is used to scale the average power of the transmitted frame in order to stay inside the spectral mask to be defined in Section 5. The Bussgang's theorem is based on Gaussian variables, but it's results are widely used, e.g., in PAPR modeling for orthogonal frequency domain multiplexing (OFDM) systems. Also in our case, the signals are not purely Gaussian, but after the pulse shape filter they are Gaussian like and we can apply Bussgang's theorem to model the non-linear limiting caused by the power amplifier model.

We have assumed a discontinuous block wise transmission where the channel is assumed to be time invariant during the transmission time of one frame. The used channel model is a modified ITU-R Vehicular A channel [13].

In Figure 2, we have presented a block diagram of our multiantenna receiver. We have extended the model provided in [4] to our SC model with FB-based frequency-domain equalizer structure, presented in [14]. The analysis FB converts the time domain signal to the frequency domain (similar to the well known DFT operation) and the synthesis FB converts the frequency domain presentation back to time domain (similar to the IDFT operation). The channel estimates are obtained in time domain after which the sub-channel wise equalization (SCE) is performed in the frequency domain with 3-tap complex FIR filter for each sub-channel. The equalizers for each diversity branch are designed based on the maximum ratio combining (MRC) criteria, presented in [15]. The channel estimates could also be obtained in the frequency domain and after suitable interpolation with DDST they could be directly used for defining the SCE equalizer tap values for each sub-channel. The FB-based receiver structure is used because it does not require a cyclic prefix (improved throughput), provides close to ideal linear equalizer performance, has good spectral containment properties (adjacent channel suppression is clearly better than with DFT based solutions) and is equally applicable also to SC-FDMA (DFT-S-OFDMA) as used in 3GPP-LTE uplink.

We assume perfect synchronization in frequency and time domain and ideal down conversion of the received signal in the *Rx* block. Several studies on DDST suitability for time and frequency synchronization have been performed, e.g., [16,17], where it has been shown that DDST is also a viable solution for low SNR synchronization. We can present the channel between transmitter and receiver as an r times oversampled discrete-time equivalent channel, $h_{\text{eq}}(n) = |h_{\text{RRC}}(t) \circ h_{\text{channel}}(t) \circ h_{\text{RRC}}(t)|_{t=nT/r} = |h_{\text{RRC}} \circ h_{\text{channel+RRC}}|_{t=nT/r}$. The nth received sample $y_i(n)$ from the ith antenna can be given as

$$
\begin{aligned}
y_i(n) = {}& \alpha\sqrt{P_{\text{AVG}}}\sum_{m=0}^{M-1} h_{\text{eq},i}(m)\breve{s}_r(n-m) \\
& + \sum_{k=0}^{K-1} h_{\text{channel+RRC},1}(k)n_G(n-k) \\
& + \sum_{l=0}^{L-1} h_{\text{RRC}}(l)w_i(n-l),
\end{aligned}
\tag{5}
$$

where M is the channel length in samples, n is the time index for r times oversampled symbol sequence, $n_G(n)$ is a noise term caused by the nonlinear amplifier, and $\breve{s}_r(n)$ is a possibly limited, oversampled transmitted symbol, which is zero if $n < 0$ or $n > rN - 1$. The noise term $w_i(n)$ is complex additive white Gaussian noise (AWGN). Because of the r times oversampling, in our case $s(k) = d(k) = p_d(k) = p_c(k) = 0$ when k modulus $r \neq 0$. The channel estimation procedures are simply repeated for each diversity branch. For this reason and for the sake of clarity, we drop out the antenna index i.

We can now rewrite the received discrete-time signal in the matrix notation as

$$
\mathbf{y} = \alpha\sqrt{P_{\text{AVG}}}\breve{\mathbf{S}}_r\mathbf{h}_{\text{eq}} + \mathbf{N}_G\mathbf{h}_{\text{channel+RRC}} + \mathbf{W}\mathbf{h}_{\text{RRC}},
\tag{6}
$$

where the matrix $\breve{\mathbf{S}}_r = \mathbf{D}_r + \mathbf{P}_{d,r} + \mathbf{P}_{c,r} + \mathbf{E}_{\text{limiter},r}$ is built from the oversampled user data symbols, data dependent pilot sequence, known cyclic pilot sequence and the additional error generated by the symbol level limiter (only with LDDST), respectively. Here $\mathbf{N}_G$ and $\mathbf{W}$ are the matrix presentations of the amplifier induced and channel induced noise terms, respectively.

Because we assume a discontinuous block-wise transmission, all matrices $\mathbf{D}_r$, $\mathbf{P}_{d,r}$, $\mathbf{P}_{c,r}$ and $\mathbf{E}_{\text{limiter},r}$ have the form

$$
\mathbf{B} = \begin{bmatrix}
b_0 & 0 & \cdots & 0 & 0 \\
b_1 & b_0 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
b_{rN_p-1} & b_{rN_p-2} & \cdots & b_1 & b_0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
b_{rN-1} & b_{rN-2} & \cdots & b_{rN-rN_p+1} & b_{N-rN_p} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 0 & b_{rN-1} \\
0 & 0 & \cdots & 0 & 0
\end{bmatrix},
\tag{7}
$$
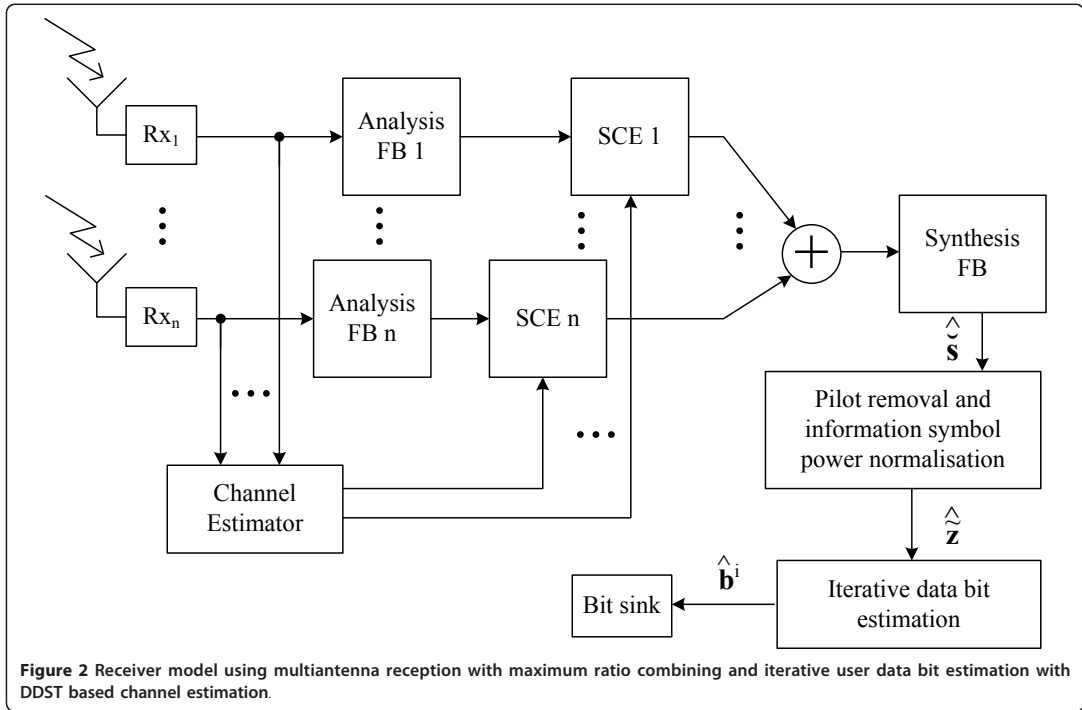
Figure 2 Receiver model using multiantenna reception with maximum ratio combining and iterative user data bit estimation with DDST based channel estimation.

including the zeros before and after the transmitted frame. Note that the oversampled matrices $\mathbf{D}_r$, $\mathbf{P}_{d,r}$, $\mathbf{P}_{c,r}$, $\mathbf{E}_{\text{limiter},r}$ are now of dimension $(rN + rN_p \times rN_p)$ and that we have assumed that $M = rN_p$. This means that in the receiver we have to do the cyclic mean calculation over $N_c + 1$ copies. Thus, the cyclic mean of the received sequence is given as

$$
\begin{aligned}
\hat{\mathbf{m}}_y &= \mathbf{J}_{Rx}\mathbf{y} \\
&= \alpha\sqrt{P_{\text{AVG}}}[\mathbf{P}_r + \hat{\mathbf{M}}_{e_{\text{limiter}},r}]\mathbf{h}_{\text{eq}} \\
&\quad + \hat{\mathbf{M}}_{nG}\mathbf{h}_{\text{channel+RRC}} + \hat{\mathbf{M}}_w\mathbf{h}_{\text{RRC}},
\end{aligned}
\tag{8}
$$

where $\mathbf{J}_{Rx} = (1/N_c)\mathbf{1}_{1\times N_c+1} \otimes \mathbf{I}_{rN_p}$. In our notation, for any vector $\mathbf{b}$, the cyclic mean vector is defined as $\hat{\mathbf{m}}_b = \mathbf{J}_{Rx}\mathbf{b} = [\hat{m}_b(0)\ \hat{m}_b(1)\ldots\hat{m}_b(rN_p - 1)]^T$, and for any matrix $\mathbf{B}$, the cyclic mean matrix is defined as

$$
\hat{\mathbf{M}}_b = \mathbf{J}_{Rx}\mathbf{B} = \begin{bmatrix} \hat{m}_b(0) & \hat{m}_b(rN_p - 1) & \cdots & \hat{m}_b(2) & \hat{m}_b(1) \\ \hat{m}_b(1) & \hat{m}_b(0) & \cdots & \hat{m}_b(3) & \hat{m}_b(2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \hat{m}_b(rN_p - 1) & \hat{m}_b(rN_p - 2) & \cdots & \hat{m}_b(1) & \hat{m}_b(0) \end{bmatrix}.
\tag{9}
$$

For example, if you set $\mathbf{b} = \mathbf{e}_{\text{limiter},r}$, then $\hat{\mathbf{M}}_{e_{\text{limiter}},r}$ is a cyclic matrix having $\hat{\mathbf{m}}_{e_{\text{limiter}},r}$ as the first column. The pilot matrix $\mathbf{P}_r$ is a cyclic matrix, having the r times

oversampled OCI pilot sequence $\mathbf{p}_r = r\mathbf{p} \otimes \mathbf{r}$ as its first column.

From the receiver frontend, the oversampled signal is provided for the channel estimator and for the analysis FB. After obtaining a channel estimate, SCE is performed in the frequency domain. More details on the equalizer structure can be found from [14,18], and references therein. After the SCE, different antenna branches are added together sub-channel wise according to the MRC principle. The composite sub-channels are then recombined in the synthesis FB, which also efficiently realizes the sampling rate reduction by 2.

After the synthesis FB, we have the Pilot removal and information symbol power normalization block. Inside this block, the received sequence power is normalized to $\sigma_{\hat{\tilde{s}}}^2 = 1 + \sigma_w^2\|\mathbf{h}_{\text{RRC}}\|^2$, which corresponds to the total received power. We have assumed that we exactly know the noise variance in the receiver. Next, we scale the power based on the pilot power allocation and remove the cyclic mean of the received sequence. If we use LDDST, we normalize the sequence based on our estimate on the average transmit power $\sigma_{\breve{s}}^2$, to be defined in (18), to obtain an estimate for the distorted data sequence,

$$\hat{\mathbf{z}} = \sigma_{\breve{\mathbf{s}}}(\mathbf{I} - \mathbf{J})\sqrt{\frac{1}{1-\gamma}}\sqrt{\frac{1 + \sigma_w^2\|\mathbf{h}_{\mathrm{RRC}}\|^2}{\sigma_{\hat{\mathbf{s}}}^2}}\hat{\mathbf{s}}. \tag{10}$$

Here $\hat{\mathbf{z}}$ is an estimate for $\mathbf{z}$ with cyclic mean set to zero and including the limiter error. Note that the cyclic mean of the limiter error is also zero.

Next, we have the Iterative data bit estimation block, where we iteratively obtain the data bit estimates. The procedures performed inside this block are described in detail in Section 6. Finally, the bit estimates are collected for bit error rate (BER) and block error rate (BLER) evaluations. The concept of (data) block in our system will be described in more detail in Section 7.

3 Symbol level limiter error modeling

Even though the earlier discussion assumed that the error caused by the symbol level limiter is purely additive, we will adopt an another model for the channel estimator modifications. In this Section, we will assume that symbol level amplitude limiter will only affect the data dependent pilot sequence, $\mathbf{p}_d$, and cyclic pilot sequence, $\mathbf{p}_c$. We model the effects by a common scaling factor and added noise. We refer to this model as the double-scaling model. We start by rewriting the limited symbol sequence as

$$\breve{\mathbf{s}} = L(\mathbf{s}) = \mathbf{d} + \beta(\mathbf{p}_d + \mathbf{p}_c) + \mathbf{n}_L. \tag{11}$$

Here the additive noise component caused by the limiter, $\mathbf{n}_L$, is assumed to be uncorrelated with $\mathbf{p}_d$ and $\mathbf{p}_c$, and it is assumed to have complex Gaussian distribution. This model is a rough approximation of the phenomena that take place in the symbol level limiter, but based on our experience it provides sufficient accuracy for the channel estimator. The main difficulty in the modeling is to incorporate the effect of the limiter on the random data-dependent pilot sequence. We have tried several models, but they all have similar or worse accuracy than the Gaussian model we are going to present here, so we chose it because of its simplicity.

We can rewrite the purely additive limiter error given in the previous Section as $\mathbf{e}_{\mathrm{limiter}} = \breve{\mathbf{s}} - \mathbf{s} = (\beta - 1)(\mathbf{p}_d + \mathbf{p}_c) + \mathbf{n}_L$. The cyclic mean of the received sequence can now be rewritten as

$$\begin{aligned}
\hat{\mathbf{m}}_y &= \mathbf{J}_{Rx}\mathbf{y} \\
&= \mathbf{J}_{Rx}\alpha\sqrt{P_{\mathrm{AVG}}}(D_r + \beta(\mathbf{P}_{d,r} + \mathbf{P}_{c,r}) + \mathbf{N}_{L,r})\mathbf{h}_{\mathrm{eq}} \\
&\quad + \mathbf{N}_G\mathbf{h}_{\mathrm{channel + RRC}} + \mathbf{W}\mathbf{h}_{\mathrm{RRC}} \\
&= \alpha\sqrt{P_{\mathrm{AVG}}}(\beta\mathbf{P}_r + (\beta - 1)\hat{\mathbf{M}}_{d,r} + \hat{\mathbf{M}}_{n_L,r})\mathbf{h}_{\mathrm{eq}} \\
&\quad + \hat{\mathbf{M}}_{n_G}\mathbf{h}_{\mathrm{channel + RRC}} + \hat{\mathbf{M}}_w\mathbf{h}_{\mathrm{RRC}}.
\end{aligned} \tag{12}$$

Because we have assumed that the limiter would affect only the pilot sequences, we have to define new methods for approximating these scaling parameters. We approximate β by generating a symbol vector consisting of all possible data symbol and pilot symbol combinations, defined as $\mathbf{s}_{\mathrm{comb},1} = \sqrt{(1-\gamma)}\mathbf{d}_l + \sqrt{\gamma}\mathbf{p}_l = 1_{N_p \times 1} \otimes \mathbf{d} + \mathbf{p} \otimes 1_{2^Q \times 1}$, where $\mathbf{d}$ is a vector containing all possible symbols, $\mathbf{p}$ is the OCI pilot sequence and Q is the number of bits per symbol. Next, we run this test sequence through the limiter and approximate the scaling factor as

$$\beta = \frac{\left|\mathbf{p}_l^H L(\mathbf{s}_{\mathrm{comb},1})\right|}{\left|\mathbf{p}_l^H \mathbf{p}_l\right|}, \tag{13}$$

where we basically calculate a correlation based weighting factor for the extended pilot sequence, $\mathbf{p}_l$. We use this same weighting factor for data dependent pilot sequence because it undergoes similar effects in the symbol level amplitude limiter.

Now the difficult question is, how can we approximate $\sigma_{e_{\mathrm{limiter}}}^2 = E[|\breve{\mathbf{s}} - \mathbf{s}|^2]$. First we have to somehow model the distribution of the cyclic mean of the transmitted sequence. The probability of a certain combination of N_c symbols follows the multinomial distribution

$$\begin{aligned}
&p(x_1, x_2, \ldots, x_k; n, p_1, p_2, \ldots, p_k) \\
&= \begin{cases} \frac{n!}{x_1! x_2! \ldots x_k!}p_1^{x_1}p_2^{x_2}\ldots p_k^{x_k}, & \text{when } \sum_{i=1}^{k}x_i = n \\ 0 & \text{otherwise}, \end{cases}
\end{aligned} \tag{14}$$

where x_i is the number of observations of a certain constellation point on a real or imaginary axis, p_i is the probability of that constellation point and in our case $n = N_c$ is the number of realizations in total per cyclic mean value. Here k is the number of constellation points per real or imaginary axis and takes the value of 2, 4 or 6 for QPSK, 16-QAM and 64-QAM, respectively. In this case, because all symbols are equally probable, $p_i = 1/k$ for all i. To get the true probability of a certain cyclic mean value, one has to add together all the probabilities of different combinations leading to that specific cyclic mean value. With high number of cyclic copies, the distribution of the cyclic mean value tends toward the Gaussian distribution, as expected based on the central limit theorem. For this reason, we have chosen to model the data dependent pilot sequence $\mathbf{p}_d$ with a continuous complex Gaussian distribution $n_{pd} \in \mathcal{N}(0, \sigma_{p_d}^2)$, where $\sigma_{p_d}^2 = E[|p_d|^2] = \sigma_d^2/N_c$, is the expected power of the data-dependent pilot sequence. In Figure 3, we have shown the true distribution of the real part of the cyclic mean component of QPSK constellation based on the multinomial distribution (which in this case

Figure 3 Example of the true distribution of the cyclic mean component based on the multinomial distribution for real part of the QPSK constellation and its Gaussian approximation with $N_c = 80$ and $\gamma = 0.1$.
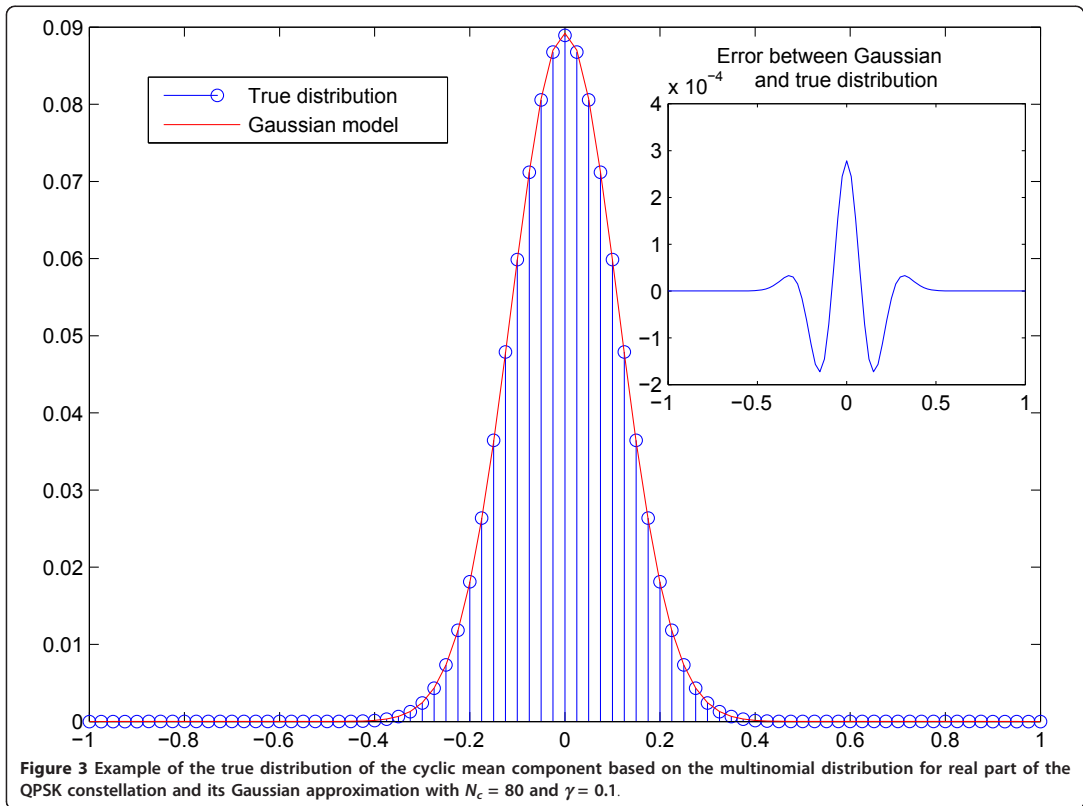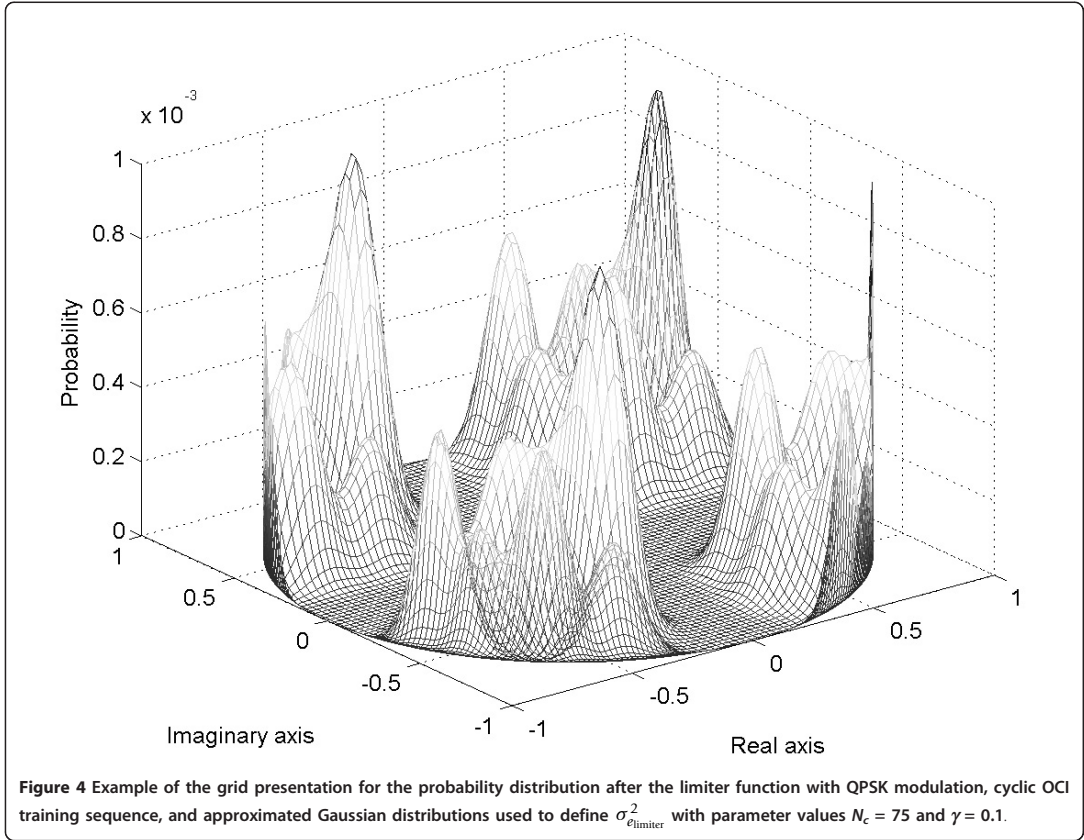
is actually binomial), its Gaussian approximation and the error between these two models. The Gaussian approximation is a good compromise for modeling purposes.

In order to approximate $\sigma^2_{e_{\text{limiter}}}$, let us first define another symbol vector consisting of all possible data symbol and pilot symbol combinations, defined as $\mathbf{s}_{\text{comb},2} = \sqrt{(1 - 1/N_c)(1 - \gamma)}\mathbf{d}_l + \sqrt{\gamma}\mathbf{p}_l$, where the power scaling factor $\sqrt{1 - 1/N_c}$ is used to ensure that the total probability over the grid model, after adding Gaussian noise modeling the cyclic mean, equals to unity. Next, we add together probability grids, in which the different grids are based on the Gaussian distribution of n_{pd} centered on a certain point of vector $\mathbf{s}_{\text{comb},2}$. The overall distribution can be given as

$P(\text{probability of symbols } \mathbf{s}_{\text{comb}} \text{ at point } x, y)$

$$= P(\mathbf{s}_{\text{comb}}, x, y) = \frac{\text{step}^2}{2^{Q N_p}} \sum_{k=1}^{2^{Q N_p}} 1/\sqrt{\pi \sigma^2_{p_d}} \tag{15}$$

$\exp\{1/\sigma^2_{p_d}[(\text{Re } (s_{\text{comb},2}(k)) - x)^2 + (\text{Im } (s_{\text{comb},2}(k)) - y)^2]\},$

where x and y present the real and imaginary axes, respectively, in a grid with values from -2 to 2. The step size used for real and imaginary axis for calculating the probabilities of cyclic mean values from the Gaussian distribution is determined by the constellation, power normalization, pilot power allocation factor and the number of cycles used in the cyclic mean calculation. For example, if we are using 16-QAM constellation with $\gamma = 0.05$ and have $N_c = 80$ cycles, the step size used is $\text{step} = 2\sqrt{1 - 0.05}/(80\sqrt{10})$, where $\sqrt{10}$ is the power normalization factor to set 16-QAM constellation average power to unity. This step now corresponds to the smallest change in the cyclic mean over possible symbols in real or imaginary axis and directly provides us a model for the discrete distribution of the cyclic mean with the defined parameters.

In Figure 4, we show as an example the generated grid model for QPSK constellation with pilot power allocation factor $\gamma = 0.1$ and number of cyclic means $N_c = 80$ after the limiter function. With QPSK the constellation power normalization factor is one, thus the step size is $\text{step} = 2\sqrt{0.9}/80$.

Figure 4 Example of the grid presentation for the probability distribution after the limiter function with QPSK modulation, cyclic OCI training sequence, and approximated Gaussian distributions used to define $\sigma^2_{e_{\mathrm{limiter}}}$ with parameter values $N_c = 75$ and $\gamma = 0.1$.

If we define $g(x, \gamma) = \sqrt{x^2 + \gamma^2}$ as a vectorized function of the distances of grid points (x, y) from the origo, we can approximate $\sigma^2_{e_{\mathrm{limiter}}}$, given as

$$\sigma^2_{e_{\mathrm{limiter}}} = \sum_{x,\gamma} \left| g(x, \gamma) - L(g(x, \gamma)) \right|^2 P(s_{\mathrm{comb}}, x, \gamma). \quad (16)$$

We will use the $\sigma^2_{e_{\mathrm{limiter}}}$ value in the ML-LMMSE channel estimator to incorporate a priori knowledge of the symbol limiter based error term.

If we now assume that $\mathbf{p}_c$, $\mathbf{p}_d$, and $\mathbf{n}_{\mathrm{limiter}}$ are uncorrelated, we can obtain the power of the limiter error with double-scaling model to be

$$\begin{aligned} \sigma^2_{n_L} &= \sigma^2_{e_{\mathrm{limiter}}} - (\beta - 1)^2 (\sigma^2_{p_d} - \sigma^2_p) \\ &= \sigma^2_{e_{\mathrm{limiter}}} - (\beta - 1)^2 (\sigma^2_d / N_c - \sigma^2_p). \end{aligned} \quad (17)$$

By using the same grid model, we can obtain our estimate of the average power of the limited symbol sequence $\sigma^2_{\tilde{s}} = E[\left| \breve{\mathbf{s}} \right|^2]$, as

$$\sigma^2_{\breve{s}} = \sum_{x,\gamma} \left| L(g(x, \gamma)) \right|^2 P(s_{\mathrm{comb}}, x, \gamma). \quad (18)$$

Here, the average power of the amplitude limited signal and the limiter error power could also be estimated by Bussgang's method [12]. However, based on our simulations, the developed model gives similar estimates and is simpler because it does not require averaging simulations for the framewise correlation calculations. Thus, it provides an alternative approach to define these parameters.

4 Channel estimation with LDDST

In this Section, we will provide the used channel estimator for LDDST. When defining the LMMSE channel estimator, we want to minimize the expected value of the squared error, $E\{|\hat{\mathbf{h}} - \mathbf{h}|^2\}$. If we now make the assumptions that the noise and the total interference experienced by the pilot sequence is AWGN, channel taps are i.i.d. and have zero mean, i.e., $E\{\mathbf{h}\} = \mathbf{0}$, the LMMSE estimator can be simplified to [19]

$$\hat{\mathbf{h}} = \left(\sigma^2 \mathbf{C}_{\hat{\mathbf{h}}_{\text{apriori}}}^{-1} + \mathbf{P}_{c,r}^H \mathbf{P}_{c,r} \right)^{-1} \mathbf{P}_{c,r}^H \mathbf{y}, \qquad (19)$$

where $\sigma^2 = \|\mathbf{h}_{\text{RRC}}\|^2 \sigma_w^2 + E[\|\mathbf{h}_{\text{channel+RRC}}\|^2] \sigma_{n_G}^2 + E[\|\mathbf{h}_{\text{eq}}\|^2] \sigma_{n_L}^2$ models the total interference power based on the Gaussian channel noise, nonlinear power amplifier caused interference and the limiter error. The channel covariance matrix, $\mathbf{C}_{\hat{\mathbf{h}}_{\text{apriori}}}$, contains the apriori information of the channel tap values. The apriori information of the channel taps is obtained through a least squares (LS) channel estimator. From (12), the LS channel estimator can be defined as

$$\hat{\mathbf{h}}_{\text{LS}} = \frac{\mathbf{P}_r^H}{\beta r^2 N_p \sigma_p^2} \hat{\mathbf{m}}_y = \left(\alpha \sqrt{P_{\text{AVG}}} - 1 \right) \mathbf{h}_{\text{eq}}$$

$$+ \frac{\alpha \sqrt{P_{\text{AVG}}} \mathbf{P}_r^H}{\beta r^2 N_p \sigma_p^2} [(1 - \beta)\hat{\mathbf{M}}_{d,r} + \hat{\mathbf{M}}_{n_L}] \mathbf{h}_{\text{eq}} \qquad (20)$$

$$+ \frac{\mathbf{P}_r^H}{\beta r^2 N_p \sigma_p^2} (\hat{\mathbf{M}}_{n_G} \mathbf{h}_{\text{channel+RRC}} + \hat{\mathbf{M}}_w \mathbf{h}_{\text{RRC}}).$$

We have assumed independent tap coefficients, which allows us to model the apriori channel correlation matrix $\mathbf{C}_{\hat{\mathbf{h}}_{\text{apriori}}}$ as a diagonal matrix. Because of the receiver pulse shape filtering, this assumption is not exactly true, but it is used to provide us simpler diagonalized LMMSE estimator model, which reduces the channel estimation complexity. We shall refer to this LMMSE estimator, that uses LS based channel estimates as a priori information, as LS-LMMSE channel estimator. The performance of the receiver could be improved with more advanced methods taking the correlation into account, like the universal basis based decomposition of the receiver pulse shape filter correlation, as was discussed in [20]. In a sense, the idea of using only the most significant components of the decomposition is similar to our idea of truncating the time window of the channel estimator to take into account only the most significant channel taps. Both methods gain in noise power reduction in the channel estimation but lose in the asymptotic accuracy.

In the channel estimator, we approximate the diagonal correlation matrix $\mathbf{C}$ by the instantaneous tap power obtained from the LS channel estimator, i.e.,

$$\mathbf{C}_{\hat{\mathbf{h}}_{\text{LS}}} = \text{diag}\left\{ \left|\hat{h}_{\text{LS}}(0)\right|^2, \left|\hat{h}_{\text{LS}}(1)\right|^2, \ldots, \left|\hat{h}_{\text{LS}}(rN_p - 1)\right|^2 \right\}. \qquad (21)$$

By assuming the cyclic OCI training sequence, the LS-LMMSE estimator can be reduced to

$$\hat{\mathbf{h}}_{\text{LS-LMMSE}} = \frac{\mathbf{P}_r^H}{\beta \left(\sigma_{\text{est}}^2 \mathbf{C}_{\hat{\mathbf{h}}_{\text{LS}}}^{-1} + r^2 N_p \sigma_p^2 \mathbf{I}_{rN_p \times rN_p} \right)} \hat{\mathbf{m}}_y. \qquad (22)$$

The variable σ_{est}^2 corresponds to the total interference power on top of each received pilot symbol and is estimated as

$$\sigma_{\text{est}}^2 = \frac{1}{\beta^2 N_c} \left[\left\|\hat{\mathbf{h}}_{\text{LS}}\right\|^2 \sigma_{n_L}^2 + (1 + 1/N_c)\sigma_w^2 \|\mathbf{h}_{\text{RRC}}\|^2 \right], (23)$$

where we do not have a term related to $\sigma_{n_G}^2$ because this value is unknown to the receiver. Similar channel estimator structure with traditional SI pilots and iterative interference canceling feedback was studied in [21].

5 PAPR analysis and spectral leakage comparison

One drawback with DDST in SC transmission is the increased PP and PAPR in the transmitted signal and spectral leakage caused by the non-linear amplifier due to the increased PAPR. These problems are well known but have received relatively little attention in the recent literature.

In a SC transmission, the PAPR of the transmitted sequence is defined after the Tx pulse-shape filter. The PP we see in the filter output depends on the maximum amplitude of the input symbols and on a portion of the absolute values of the filter coefficients, depending on the oversampling. Because we have fixed the Tx pulse-shape filter, only the maximum amplitudes of the input symbols effect the observed PAPR.

There are two main reasons for increased symbol level amplitude in DDST. First of all, we increase the amplitude range related to a certain constellation by adding a power scaled pilot sequence on top of a power scaled symbol sequence. The second main reason for increased amplitude is the possibility of a cyclic mean (data dependent pilot) component with relatively high amplitude. When this component is added on top of data and known pilot symbols, and if the angles of these complex variables happen to align, then the total symbol amplitude is significantly increased.

In this Section, we will first discuss the worst case PP and PAPR effects in more detail and after that we will describe the reference spectral power mask and related simulations and results.

5.1 PAPR analysis and simulated results

For the analysis and results in this section, we have used oversampling ratio equal to four, $r = 4$. The worst case evaluations are based on the filter taps with separation of r samples that have the highest sum-power. This is because the transmitted symbol sequence is oversampled by factor r, so then for each output only every rth filter tap value participates in the corresponding power value. In other words, the filter model used in the following derivations is defined as $h_{\text{RRC}}(\mathbf{i})$, where the set of indices

i is chosen based on criteria

$$\left\{\mathbf{i} = [k, k + r, \ldots, k + \mathrm{nr}] \mid \max_k \left[\left(\sum_{i \in \mathbf{i}} |h_{\mathrm{RRC},Tx}(i)| \right)^2 \right] \right\}, \quad (24)$$

where $k \in [0,1,\ldots,r-1]$ and $k + \mathrm{nr} \leq N_{\mathrm{RRC}}$. With RRC transmit pulse shape filter of degree 64 and $r = 4$, the starting index which maximizes the sum-power is $k = 2$. Because the RRC filter acts also as a oversampling filter, the taps of the filter are multiplied by the oversampling factor r in order to keep the average transmitted power equal to unity.

First, we define the worst case symbol level PP. Assume now that $d(k) = ae^{j\varphi}$ is some corner symbol with amplitude a and all the other symbols present in the cyclic mean calculation, $d(k + iN_p) = ae^{j(\varphi - \pi)}$ with $i = 1,2,\ldots,N_c - 1$, are opposite corner symbols with amplitude a. Then the data dependent pilot added on top of $d(k)$ is equal to

$$\begin{aligned}
p_d(k) &= -\frac{1}{N_c} \sum_{i=0}^{N_c-1} d(k + iN_p) \\
&= -\frac{1}{N_c} [(N_c - 1)(ae^{j(\varphi - \pi)}) + ae^{j\varphi}] \quad (25) \\
&= \frac{(N_c - 2)}{N_c} ae^{j\phi} \\
&= \frac{N_c - 2}{N_c} \sqrt{1 - \gamma} a_{\max} e^{j\phi},
\end{aligned}$$

which corresponds to the worst case peak amplitude with the data dependent pilot sequence and its value depends on the used constellation and the pilot power allocation factor γ. The worst case symbol level PP is defined for an aligned pilot $p_c(k)$ which has amplitude $\sqrt{\gamma}$. By aligned, we mean that the arguments of data and the pilot are equal, $\angle d(k) = \angle p_c(k) = \varphi$. Now we can write the worst case symbol level PP as

$$\begin{aligned}
\mathrm{WPP}_s &= |d(k) + p_d(k) + p_c(k)|^2 \\
&= \left[\left(1 + \frac{N_c - 2}{N_c}\right) \sqrt{1 - \gamma} a_{\max} + \sqrt{\gamma} \right]^2. \quad (26)
\end{aligned}$$

By using (26), we can define then the worst case PP after the transmit pulse shape filtering to be

$$\begin{aligned}
\mathrm{WPP}_{Tx,\mathrm{DDST}} &= \left(\sum_{i \in \mathbf{i}} |h_{\mathrm{RRC}}(i)| \right)^2 \\
&\quad \left[\left(1 + \frac{N_c - 2}{N_c}\right) \sqrt{1 - \gamma} a_{\max} + \sqrt{\gamma} \right]^2, \quad (27)
\end{aligned}$$

For TDMT, the worst case PP after the transmit pulse shape filtering is

$$\mathrm{WPP}_{Tx,\mathrm{TDMT}} = a_{\max}^2 \left(\sum_{i \in \mathbf{i}} |h_{\mathrm{RRC}}(i)| \right)^2. \quad (28)$$

If we use the presented hard symbol level limiter in the transmitter, then the worst case symbol level PP can be given as

$$\mathrm{WPP}_{s,\mathrm{limited}} = |L(d(k) + p_d(k) + p_c(k))|^2 = a_{\max}^2, \quad (29)$$

which is the same as with TDMT. Then the worst case PP after the RRC filtering is

$$\mathrm{WPP}_{Tx,\mathrm{DDST,limited}} = a_{\max}^2 \left(\sum_{i \in \mathbf{i}} |h_{\mathrm{RRC}}(i)| \right)^2. \quad (30)$$

which is equal to TDMT case.

With the PPs defined, we can define the PAPRs for different cases. While reading the results for PAPR from Table 1, one should note the difference in the average powers used to define these PAPR results. The average power of a TDMT signal is given as $E[|s_{\mathrm{TDM}}|^2] = 1$. For DDST based system, the average power of the signal is $E[|s|^2] = (1 - 1/N_c)\sigma_d^2 + \sigma_p^2$. The weighting factor $(1 - 1/N_c)$ is caused by the removal of the cyclic mean from the data sequence. Now the worst case PAPR for DDST without limiter before and after the transmitter pulse shape filter can be given as

$$\begin{aligned}
\mathrm{WPAPR}_s &= \frac{\mathrm{WPP}_s}{E[|s|^2]} \\
&= \frac{\left[\left(1 + \frac{N_c - 2}{N_c}\right) \sqrt{1 - \gamma} a_{\max} + \sqrt{\gamma} \right]^2}{(1 - 1/N_c)\sigma_d^2 + \sigma_p^2}, \quad (31)
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{WPAPR}_{Tx,\mathrm{DDST}} &= \frac{\mathrm{WPP}_{Tx,\mathrm{DDST}}}{E[|s|^2]} \\
&= \frac{\left(\sum_{i \in \mathbf{i}} |h_{\mathrm{RRC}}(i)| \right)^2 \left[\left(1 + \frac{N_c - 2}{N_c}\right) \sqrt{1 - \gamma} a_{\max} + \sqrt{\gamma} \right]^2}{(1 - 1/N_c)\sigma_d^2 + \sigma_p^2}. \quad (32)
\end{aligned}$$

The average power for LDDST is given as $E[|\breve{s}|^2] = \sigma_{\breve{s}}^2$ and is defined based on the Gaussian grid model in (18) in Section 3. The PAPRs for the limited case can be written as

$$\mathrm{WPAPR}_{s,\mathrm{limited}} = \frac{\mathrm{WPP}_{s,\mathrm{limited}}}{E[|\breve{s}|^2]} = \frac{a_{\max}^2}{\sigma_{\breve{s}}^2}, \quad (33)$$

and

$$\mathrm{WPAPR}_{Tx,\mathrm{DDST,limited}} = \frac{a_{\max}^2 \left(\sum_{i \in \mathbf{i}} |h_{\mathrm{RRC}}(i)| \right)^2}{\sigma_{\breve{s}}^2}. \quad (34)$$

Finally, the PAPR for the TDMT case equals

$$
\begin{aligned}
\text{WPAPR}_{Tx,\text{TDMT}} &= \frac{\text{WPP}_{Tx,\text{TDM}}}{E[|s_{\text{TDM}}|^2]} \\
&= a_{\max}^2 \left(\sum_{i \in \mathbf{i}} |h_{\text{RRC}}(i)| \right)^2.
\end{aligned}
\tag{35}
$$

In Table 1, we have calculated different symbol level and transmitted signal related worst case PPs and PAPRs for different constellations with pilot power allocation factor $\gamma = 0.1$. As we can see, the hard limiter significantly decreases the worst case PPs and PAPRs and the limited worst case PAPRs are close to the TDMT cases, as was desired.

If we assume that with DDST we want to set the PP at the transmit pulse shape filter output to be at a similar level as with TDMT, based on Table 1, a significant back-off is required. With symbol level amplitude limiter we can remove this backoff requirement. As a downside, the amplitude limiter causes additional interference in the transmitted symbols, which might be significant especially with higher order modulations.

In Table 2, the different simulated PPs and PAPRs are given for each constellation. The simulated values were obtained by finding the maximum PAPR over 100,000 random frame realizations. These results provide more insight on the average PAPR performance of the given system with different training methods, and show that the defined analytic worst case PPs and PAPRs are reliable upper bounds.

As expected, the PP and PAPR results with DDST are not as bad as the worst case studies suggested. The main benefit of using symbol level limiter seems to be with QPSK and 16-QAM constellations, where significant reduction in PAPR can be achieved. 64-QAM has quite similar performance with and without symbol level limiter. In Figure 5, an example of the complementary cumulative distribution functions (CCDF) for PP and PAPR distributions with QPSK constellation are shown.

Table 1 WPP and WPAPR for the used constellations with parameter values $N_c = 75$, $N_p = 60$, and $\gamma = 0.1$

	QPSK	16-QAM	64-QAM
WPP$_s$ (26)	4.8	8.0	10
WPP$_{s,\text{limited}}$ (29)	1	1.8	2.3
WPP$_{Tx,\text{DDST}}$ (27)	25.6	42.7	53.8
WPP$_{Tx,\text{LDDST}}$ (30)	5.3	9.6	12.5
WPP$_{Tx,\text{TDMT}}$ (28)	5.3	9.6	12.5
WPAPR$_{Tx,\text{DDST}}$ (32)	25.9	43.3	54.6
WPAPR$_{Tx,\text{LDDST}}$ (34)	5.4	10.2	12.7
WPAPR$_{Tx,\text{TDMT}}$ (35)	5.3	9.6	12.5

All values are given in linear scale

Table 2 Simulated PPs and PAPRs for the used constellations with parameter values $N_c = 75$, $N_p = 60$, and $\gamma = 0.1$

	QPSK	16-QAM	64-QAM
PP$_s$	2.8	3.9	4.6
PP$_{s,\text{limited}}$	1	1.8	2.3
PP$_{Tx,\text{DDST}}$	6.6	8.7	9.3
PP$_{Tx,\text{LDDST}}$	4.7	7.6	8.9
PP$_{Tx,\text{TDMT}}$	5.3	7.7	9.1
PAPR$_{Tx,\text{DDST}}$	6.8	9.0	9.5
PAPR$_{Tx,\text{LDDST}}$	5.9	8.2	9.2
PAPR$_{Tx,\text{TDMT}}$	5.3	7.8	9.2

All values are given in linear scale

Here we can see that the PAPR distributions are similar but the PP distributions are quite different.

5.2 Spectral leakage with SSPA amplifier model

In this section we will study the spectral re-growth with different training methods and with QPSK, 16-QAM, and 64-QAM constellations. The power amplifier model was given in Section 2. We have chosen to use values $\nu = 1$ and $p = 3$ for the simulations. Because we have assumed that the power amplifier is matched to work with TDMT transmission, we have set the 1 dB compression point of the power amplifier based on the 64-QAM constellation PP distribution. The chosen amplitude limit is related to the PP which gives us 1% probability in the CCDF. Thus, from the results obtained in the previous section, we can look for the PP with 64-QAM that $P(\text{PP}_{64-QAM} \leq P_{1\text{dB}}) = 0.01$. Based on our simulations, this value is equal to $P_{1\text{dB}} = 4.8$ dB. Now, we use this power value to solve the power amplifier saturation amplitude. The amplitude corresponding to the 1 dB compression point is $A = 10^{4.8/20}$ and the saturation amplitude can be solved to be

$$
A_0 = \nu A \left(10^{p/10} - 1 \right)^{\frac{-10}{2p}},
\tag{36}
$$

which gives us $A_0 \approx 1.739$.

The used spectral mask is based on 3GPP technical specification for E-UTRA user equipment [22]. The used required attenuation levels are based on 23 dBm transmission power in the used 20 MHz bandwidth and Table 6.6.2.2.2-1 in page 44 of [22]. We chose the values of this Table because it provides the most strict attenuation mask. The obtained attenuation levels are given in Table 3 with respect to the distance from the channel band edge. This distance is defined as an out-of-band frequency distance, Δf_{OOB}. The required attenuation levels are defined for a measurement bandwidth of 1 MHz.

For the simulations, we have assumed to use 20 MHz channel bandwidth, 18 MHz symbol frequency and a roll-off factor 0.1 in the RRC filter. We wanted to keep
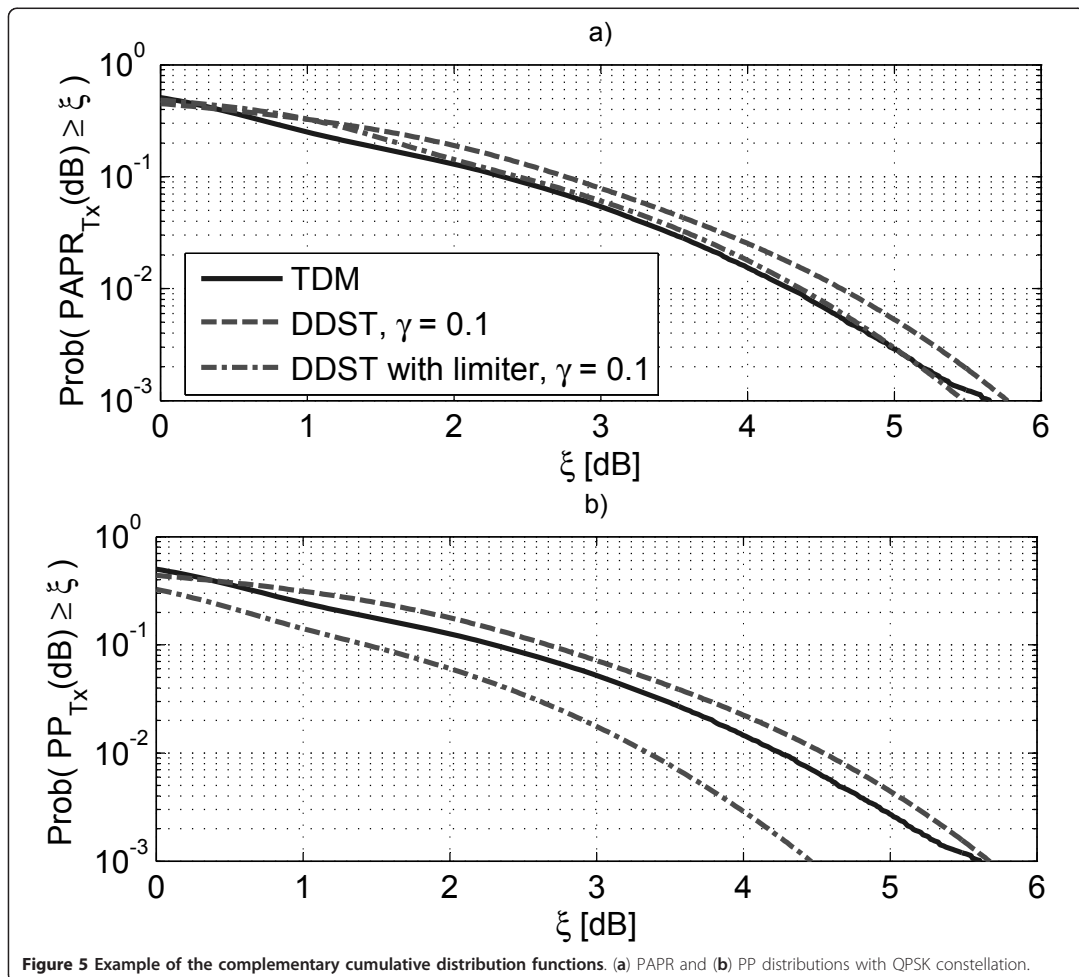
Figure 5 Example of the complementary cumulative distribution functions. (**a**) PAPR and (**b**) PP distributions with QPSK constellation.

the roll-off factor small because we are aiming toward very high spectral efficiency. For different training methods and constellations, we ran the simulations looking for smallest IBO with 0.5 dB step in the average transmitted power, P_{AVG}. We have defined the input backoff (IBO) as IBO $= 10\log_{10}(A_0^2/P_{\mathrm{AVG}})$. Based on the results, we chose the smallest IBO for each training method and constellation which leads to spectral leakage that stays below the given spectral mask. The obtained IBO and output backoff (OBO) results are provided in the Table 4. The OBO is defined as the maximum output power to the average output power ratio, given as OBO $= 10\log_{10}(A_0^2/E[G(\mathbf{x})^2])$.

As expected, based on the PP and PAPR analysis, we can reach significantly lower OBO when using limited

DDST with QPSK constellation. With 16-QAM constellation we can decrease the OBO somewhat with symbol level limiter. With 64-QAM, meaningful gains were not achieved with symbol level amplitude limiter. These IBO values are used in Section 7 when we compare the throughput performance of different training methods.

Next, we will return to the actual implementation of the iterative receiver used with limited DDST before we study the throughput performance with different training methods.

6 Iterative receiver algorithms

The receiver operations before the iterative data bit estimation were already described in Section 2. In this section we discuss in more detail the operations performed

Table 3 Attenuation at distance Δf_{OOB} from the channel band edge

Δf_{OOB} [MHz]	Attenuation requirement [dB]
±0-1	-15.76
±1-5.5	-22.99
±5.5-25	-34.99

inside the iterative data bit estimation block, shown in more detail in Figure 6.

We have used notation $\hat{\bar{\mathbf{z}}}$ to represent our estimates of the data symbol sequence, including the limiter error, with cyclic mean set to zero, obtained from the pilot removal and information symbol power normalization block, as shown in Figure 2. We use $\hat{\bar{\mathbf{z}}}$ as a initial data symbol estimates to generate hard symbol based cyclic mean estimate in the hard symbol based $\mathbf{p}_d$ estimation and compensation block. Inside this block, we generate hard symbol estimates based on $\hat{\bar{\mathbf{z}}}$, calculate their cyclic mean and add it to $\hat{\bar{\mathbf{z}}}$, to obtain initial symbol estimates $\hat{\mathbf{d}}^0$. Here superscript 0 points out that these symbol estimates are obtained before coded feedback. This idea was presented in [4], and we use it before the first soft symbols to bits mapping.

We start the iterative reception process by using $\hat{\mathbf{d}}^0$ to generate soft coded bit estimates $\hat{\bar{\mathbf{b}}}$ in the *soft symbols-to-bits* block. These are then provided to the soft-input soft-output (SISO) decoder from which we obtain our first soft decoded bit estimates to be provided for the $\mathbf{p}_d$ and $\mathbf{e}_{\text{limiter}}$ estimation and compensation block and for bit error evaluation. This block is presented in more detail in Figure 7, where superscript i refers to the iteration number. These procedures, before we obtain the first feedback data symbol estimates, $\hat{\mathbf{d}}^1$, are considered to happen in the zeroth feedback iteration ($i = 0$). In our notation, after first pass through channel decoder, symbol estimation and compensation processes, we obtain our first feedback data symbol estimates $\hat{\mathbf{d}}^1$, to be used for soft bit estimation.

Table 4 Simulation based IBO and OBO results for different training methods and constellations

Training method/constellation	QPSK	16-QAM	164-QAM
Required IBO [dB]			
TDMT	5.3	5.8	5.8
DDST	5.3	5.8	5.8
DDST with limiter	3.8	5.3	5.8
Corresponding OBO [dB]			
TDMT	5.5	6.0	6.0
DDST	5.6	6.1	6.1
DDST with limiter	5.0	5.8	6.1

All values are given in decibels [dB]

The operations inside the $\mathbf{p}_d$ and $\mathbf{e}_{\text{limiter}}$ estimation and compensation block, shown in Figure 7, are performed as follows. First we generate soft symbol estimates based on the latest soft bit estimates $\hat{\mathbf{b}}^i$, which are equal to the log-likelihood presentation of the a posteriori probabilities obtained from the soft decoder. The soft symbols are given by equation

$$\hat{d}_v^i = \sum_{a=1}^{|A|} d_a p\left(d_a | \hat{\mathbf{b}}_v^i\right), \quad 0 \leq v \geq N - 1, \tag{37}$$

where $|A|$ gives the number of symbols in alphabet A, v is a symbol index, $\hat{\mathbf{b}}_v^i$ are the soft bit estimates related to the vth symbol, and $p\left(d_a | \hat{\mathbf{b}}_v^i\right)$ is the probability of a symbol d_a, given the latest soft bit estimates $\hat{\mathbf{b}}_v^i$. The probability of a symbol d_a is defined as

$$p\left(d_a | \hat{\mathbf{b}}_v^i\right) = 2^{-Q} \prod_{q=1}^Q \left[1 + \bar{b}_{d_a}(q) \tanh\left(\frac{\hat{b}_v^i(q)}{2}\right)\right], \tag{38}$$

where Q is the number of bits per symbol, $\bar{b}_{d_a}(q) \in [-1, +1]$ is the qth bit of the hypothesis d_a, and $\hat{b}_v^i(q)$ is the log-likelihood presentation of the a posteriori probability related to the qth bit of the vth symbol in the ith iteration, given as

$$\hat{b}_v^i(q) = \log\left(\frac{P_{app}(b_v^i(q) = 1)}{P_{app}(b_v^i(q) = 0)}\right). \tag{39}$$

We have also normalized the variance of the soft symbol vector, $\hat{\mathbf{d}}^i$, to be equal to unity. This improves the feedback performance when the soft bit estimates have very low reliability. In our simulations, using soft symbol feedback for the limiter error estimation provided better results than using hard symbol feedback.

Then, we calculate the symbol wise cyclic mean and remove it from the symbol sequence to obtain $\hat{\mathbf{z}}^i$. Now $-\hat{\mathbf{p}}_d^i$ is an improved estimate of the cyclic mean, assuming that the SISO decoder has been able to reduce the number of bit errors in the detected bit sequence. Next, we add the known pilot sequence on top of the sequence $\hat{\mathbf{z}}^i$ to get $\hat{\mathbf{s}}^i$ and provide this sequence to the amplitude limiter. Then we calculate the limiter error estimate based on the input and the output of the limiter function and an improved estimate of the average power, $\sigma^2_{\hat{\bar{s}}^i}$. At this point, when $i > 0$, we obtain our first estimate of the limiter error. Based on our results, it is better to estimate the limiter error after the channel
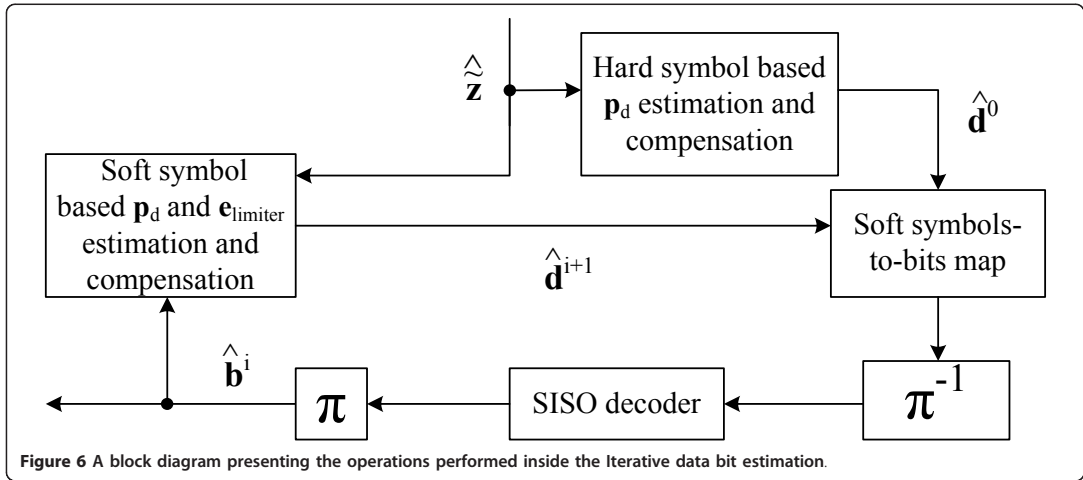
Figure 6 A block diagram presenting the operations performed inside the Iterative data bit estimation.

decoder and not based on the uncoded hard symbol estimates $\hat{\mathbf{d}}^0$. With low code rates (low E_b/N_0 region) the uncoded limiter error estimation leads to worse performance in all iterations. Then again, with high code rates (high E_b/N_0 region) uncoded limiter error estimation improves the BLER performance at the 0th iteration, but the iterative gain decreases, leading to worse performance at the fifth iteration.

Based on this improved average amplitude estimate, we can obtain improved symbol estimates by rescaling the average power of the received sequence, remembering that we have already scaled the incoming sequence by $\sigma_{\breve{s}}$ in (10). Finally, we can generate new symbol

estimates by adding to the received symbol estimates $\hat{\breve{\mathbf{z}}}$ the latest cyclic mean and limiter error estimates, given as

$$
\begin{aligned}
\hat{\mathbf{d}}^{i+1} &= \frac{\sigma_{\breve{s}}^i}{\sigma_{\breve{s}}}\hat{\breve{\mathbf{z}}} - \hat{\tilde{\mathbf{e}}}_{\text{limiter}}^i - \hat{\mathbf{p}}_d^i \\
&= \frac{\sigma_{\breve{s}}^i}{\sigma_{\breve{s}}}\hat{\breve{\mathbf{z}}} - (\mathbf{I} - \mathbf{J}_{Tx})\hat{\mathbf{e}}_{\text{limiter}}^i + \mathbf{J}_{Tx}\hat{\mathbf{d}}^i.
\end{aligned}
\tag{40}
$$

We remove the cyclic mean of the estimated limiter error $\hat{\mathbf{e}}_{\text{limiter}}^i$, because we have completely removed the cyclic mean from $\hat{\breve{\mathbf{z}}}$, including the limiter error.
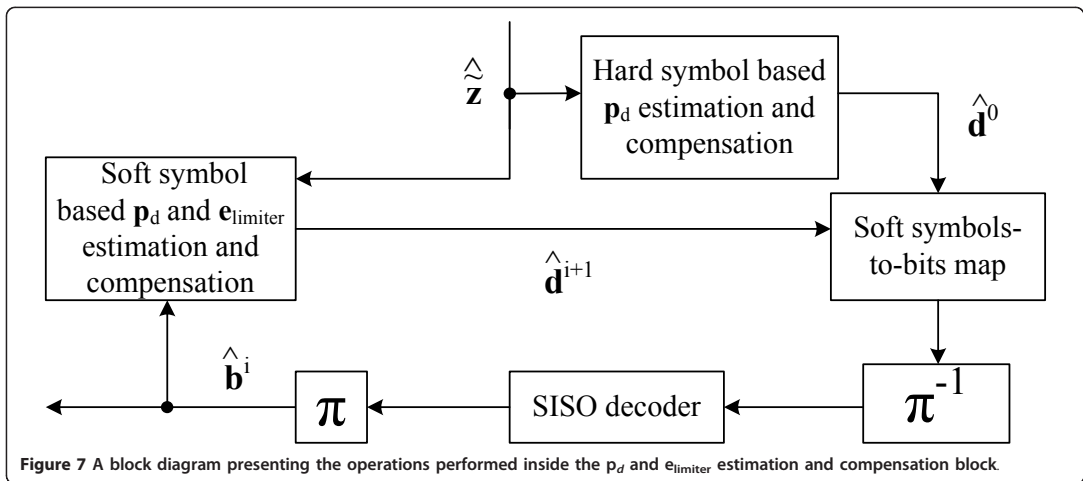

Figure 7 A block diagram presenting the operations performed inside the p_d and e_{limiter} estimation and compensation block.

Based on our results, it is better not to use the extrinsic information obtained from the channel decoder as a priori information in the soft symbols-to-bits mapping, if this information is already used to improve the cyclic mean estimate. This is probably because we are using the same information twice inside the same loop, thus losing the independence of the a priori information. We can use it as a priori information if we do not improve the cyclic mean, but based on our studies this does not provide as good iterative gain in the receiver. This could be because of the error averaging nature of the cyclic mean computation.

Here we remind the reader, that even without symbol level amplitude limiter, we have to use iterative detection algorithm for the cyclic mean estimation. Of course, the limiter error estimation is not required. Therefore, in the simulation results presented in Section 7, the throughput results obtained with DDST also include five feedback iterations.

For a reader interested in a pure SI training with iterative reception, a good starting point is, for example, [23]. In this article a computationally efficient, iterative frequency-domain equalization and channel estimation is presented. In this article, we have not considered of including the channel estimation process in the iterative loop because with DDST there is no interference from the data symbols to the known pilot symbols. Nonetheless, when there is symbol level limiter involved, we could feedback the cyclic mean of the limiter error estimate in order to improve the channel estimates with LDDST. In addition, in SISO case or in spatially multiplexed MIMO case, the feedback filtering used also in [23], is of great interest and provides interesting topics for future research.

7 Performance comparisons

In this section, we will first provide some results demonstrating the performance of our iterative receiver algorithm. In the end, spectral efficiency comparisons between TDMT and DDST based training are provided. This is, after all, the most important topic of this article. We will investigate whether the end user spectral efficiency is really improved with DDST and do we gain something by using a symbol level amplitude limiter.

The used channel model is a block-fading extended ITU-R Vehicular A channel with approximately 20 MHz bandwidth [13]. The maximum delay spread of the channel is 78 samples. In [13], the channel model was defined for sampling interval t_s = 32.55 ns where as in our system the sampling interval is t_s = 27.78 ns. This modification has a minor effect on the spectral correlation properties of the channel. However, the main idea is only to do some initial comparisons in the possible throughput performance between DDST and TDMT training based

systems. Therefore, the used model provides a good starting point for the simulations.

The oversampling in the receiver allows us to efficiently realize the RRC filtering in frequency domain in combination with the channel equalization process. More details can be found in [14] and references therein. In this article we have considered single-input single-output (SISO), and 1 × 2 and 1 × 4 single-input multiple-output (SIMO) antenna configurations with MRC equalizer.

In our simulations, the channel estimator length is rN_p = 120 while the true equivalent channel length, including the effects of transmitter and receiver RRC filters, is $N_{channel}$ + 1 + $2N_{RRC}$ = 206 samples. This kind of short channel estimator was studied in [21,24]. The reason behind using short channel estimator is to maximize the number of cycles, N_c, with the cost of minimizing the estimator length, N_p. Because we are estimating the equivalent channel, we can ignore channel tap values close to zero, which are caused by the heavy tailing of the RRC filters. In the presented simulations we have used values N_c = 75 and N_p = 60 with DDST and LDDST. This gives us a good compromise with the estimator accuracy and achievable number of cyclic copies. Especially with QPSK modulation, when we are working in a high noise environment, it is worth to consider sacrificing the channel estimation accuracy to achieve better noise power averaging through increased number of cyclic copies. With higher order constellations, in addition to the improved noise averaging, with increased number of copies we can also decrease the variance of the data dependent training sequence, $\mathbf{p}_d$, and this improves the accuracy of the first symbol estimates.

The channel codec uses turbo code [25] with generator matrix $G = [1\frac{15}{13}]$. We have used the max-log-MAP algorithm presented in [26] without any correction factor for the max-operator. The extrinsic information exchanged between the component decoders is weighted by a factor 0.75 to reduce the error propagation, as proposed in [27]. Iterations in the turbo decoder are terminated based on the hard-data-aided algorithm presented in [28]. The used interleavers are bitwise S-interleavers [29], where the distance parameter is defined as $S = \sqrt{U/2}$ where U is the length of the unit which is interleaved. In channel interleaving the unit is the whole transmitted frame $U = QN$, where Q is the number of bits per symbol and N is the number of symbols per transmitted frame. We divide each transmitted frame into Q coded blocks. Inside the turbo codec the length of the interleaved unit is equal to one uncoded data block $U = \lfloor R(N - 2m) \rfloor$, where m = 3 is the memory length of the component encoder and the term 2 m is caused by the unpunctured termination bits [30].

We have run the simulations for QPSK, 16-QAM, and 64-QAM constellations with code rates $R = 0.5$, $R = 0.67$ and $R = 0.75$. With TDMT pilots, the number of transmitted data symbols in each frame is decreased by the number of pilot symbols, which is set to be 450 in our simulations (10% of the frame duration). The TDMT pilots are the first 450 binary symbols from a Gold code of length 512 symbols [31] with unity power. The channel estimator length is equal to the equivalent channel length. With DDST, we decided to provide same portion of total power for the pilots, thus $\gamma = 0.1$. This gives us a fair comparison between TDMT training and DDST based transmission, because the channel estimation MSE of basic least-squares channel estimator with DDST is the same as with TDMT, if equal amount of power is allocated for the pilots [4]. The optimization of the pilot powers with TDMT or DDST for channel estimation with transmitted average power and PP restrictions is an interesting and open problem, but is out of the scope of this article. Some additional simulation parameters related to the simulation model are given in Table 5.

In all the simulated cases we have used the maximum of five feedback iterations for $\hat{\mathbf{p}}_d$ and $\hat{e}_{\text{limiter}}$ estimation. Typically, for QPSK modulation two and for 16-QAM modulation three feedback iterations already provide relatively good performance. With 64-QAM modulation we need five feedback iterations to ensure convergence in all of the cases. Example of the typical BLER behavior over iterations with LDDST using amplitude limiter with different constellations, compared to TDMT, is shown in Figure 8. We have assumed that the receiver does not know the IBO used in the transmitter and this degrades the performance results in all of the simulated cases.

One rather intriguing problem while planning the spectral efficiency comparison was the choice of the reference power. The comparison of performance with DDST and TDMT based systems is not so trivial and one has to be

careful about what to compare and how these results should be interpreted.

In the simulations, we chose to do the performance comparisons with respect to the energy per transmitted data bit over one sided noise spectral density, E_b/N_0. We have chosen this parameter because what matters most in modern wireless communications is the used energy per data bit to transmit with certain spectral efficiency. We have defined the SNR based on E_b/N_0 as

$$\text{SNR} = \frac{E_b Q R_{\text{true}}}{N_0 r}, \qquad (41)$$

where Q is the number of bits per symbol, R_{true} is the true coding rate (including the effect of possible termination bits, block length modifications with zero padding, etc.), and $r = 2$ is the oversampling rate used in the receiver.

Figures 9 and 10 present spectral efficiency results for DDST, LDDST and for TDMT training, using also a LS-LMMSE type equalizer, with QPSK modulation and with 16-QAM and 64-QAM modulations, respectively. From Figure 9 we can observe how the increased average transmit power allowed by the symbol level amplitude limiter improves the spectral efficiency in the low E_b/N_0 range with QPSK modulation. In Figure 10 we have shown the performance with higher order modulations. Here, the performance of LDDST compared to DDST is quite similar. Clearly, both DDST based systems improves the spectral efficiency over the whole E_b/N_0 range for each antenna configuration. The maximum spectral efficiency difference for each constellation is equal to 10%, which corresponds to the pilot overhead of TDMT.
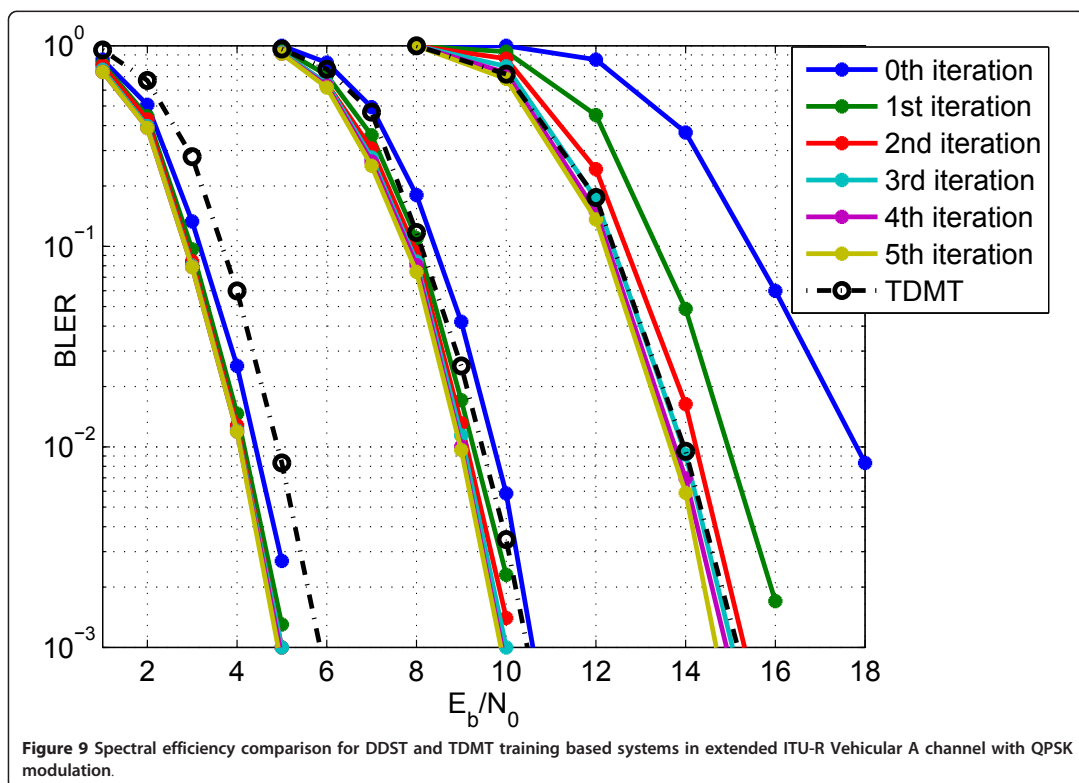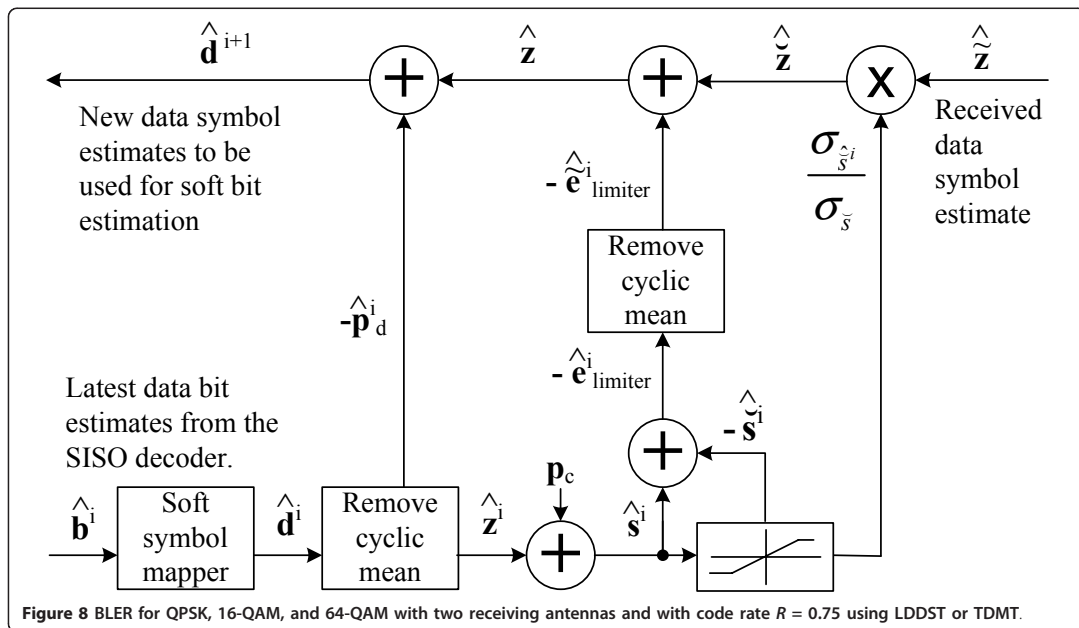
With the proposed symbol level amplitude limiter we can obtain improved spectral efficiency performance with QPSK modulation in all antenna configurations. With 16-QAM or 64-QAM modulations, LDDST and DDST have quite the same performance. Possibly, one could improve the LDDST performance with higher order modulations by tighter limiting bounds. In addition, by first performing tighter limiting and after that removing the cyclic mean, we could decrease the limiter error effect in the channel estimation and possibly improve the system performance. These topics are left for future studies.
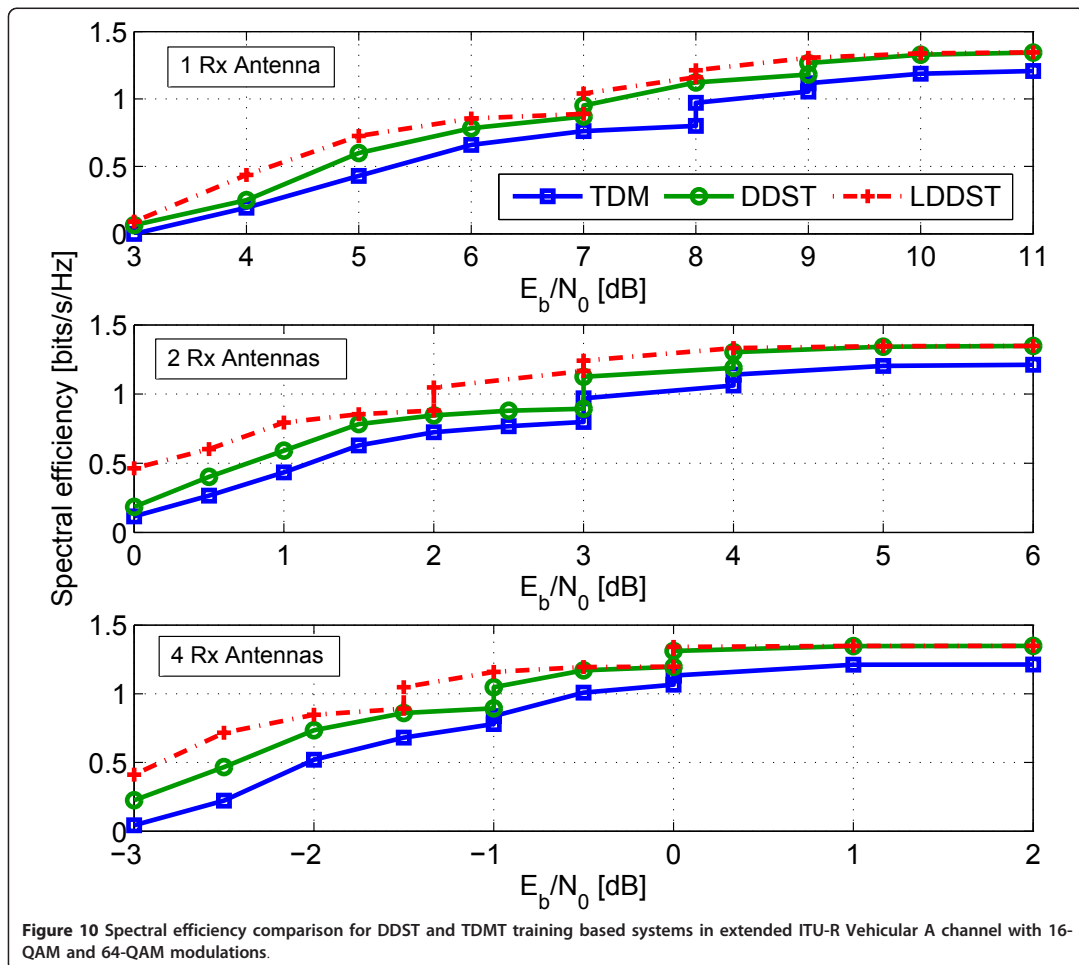
8 Conclusion

In this article, we have discussed the effects of a DDST based training on the signal PP and PAPR distributions. We demonstrated that the PP and PAPR distributions of the DDST based training have longer tails and therefore there is a higher probability for big PAPR values. Especially, with constant amplitude modulations like QPSK, the average PAPR is significantly increased. Furthermore, the effects of the increased PAPR on the spectral leakage

Table 5 Simulation parameters

Symbol rate	18 MHz
Signal bandwidth	19.8 MHz
Frame duration	250 μs
Order of the RRC filter	64
RRC roll-off	0.1
Symbols per frame	4,500
TDMT pilot symbols per frame	450
Number of feedback iterations	5
No. of subbands in the analysis bank	1,024
No. of subbands in the synthesis bank	512
FB Overlapping factor	5
FB roll-off	1

Figure 8 BLER for QPSK, 16-QAM, and 64-QAM with two receiving antennas and with code rate $R = 0.75$ using LDDST or TDMT.

Figure 9 Spectral efficiency comparison for DDST and TDMT training based systems in extended ITU-R Vehicular A channel with QPSK modulation.

Figure 10 Spectral efficiency comparison for DDST and TDMT training based systems in extended ITU-R Vehicular A channel with 16-QAM and 64-QAM modulations.

with SSPA amplifier model were studied. It was shown, that DDST does not require higher IBO compared to TDMT, but does provide slightly worse OBO performance. The proposed symbol level limiter can decrease further the IBO and OBO requirements with QPSK and 16-QAM constellations. The reduced OBO and IBO may significantly ease the design, implementation and cost of the required power amplifier. With QPSK modulation the symbol level limiter also clearly decreases the spectral re-growth and improves the spectral efficiency performance via higher average transmitted power.

Based on our results, with QPSK and 16-QAM, one should consider using LDDST to allow higher average transmitted power (lower OBO) and to achieve improved throughput compared to DDST. With higher order constellations symbol level amplitude limiter, as

presented in this article, doesn't seem to provide significant benefit.

With DDST, with or without symbol level amplitude limiter, the complexity increase compared to traditional TDMT training can be approximated by the complexity of the SISO decoder used. In the soft feedback loop with DDST, with or without symbol level amplitude limiter, the SISO decoder is dominating the detection complexity. Thus, the average increase in the detection complexity compared to TDMT, is roughly the average number of feedback iterations times the number of blocks decoded in average in each feedback iteration times the average complexity of decoding one block in the SISO decoder. With TDMT no feedback iterations are required.

The performance comparisons between DDST and TDMT based system showed that DDST can provide

similar or better performance over the whole E_b/N_0 range with all antenna configurations. The proposed symbol level amplitude limiter improves the throughput performance of the DDST in the low E_b/N_0 range with all antenna configurations tested.

In addition to careful performance analysis and comparisons, we have provided some new ideas for PAPR control with DDST, for modeling the effects of symbol level limiter in channel estimation, and for modeling the cyclic mean distribution based on multinomial distribution or its Gaussian approximation.

Competing interests
The authors declare that they have no competing interests.

References
1. P Hoeher, F Tufvesson, Channel estimation with superimposed pilot sequence, in *Proc IEEE Global Telecommunications Conference 1999 GLOBECOM '99*, vol. 4. (Janeireo, Brazil, 1999), pp. 2162–2166
2. AG Orozco-Lugo, MM Lara, DC McLernon, Channel estimation using implicit training. IEEE Trans Signal Process. 52(1), 240–254 (2004). doi:10.1109/TSP.2003.819993
3. SAK Jagannatham, BD Rao, Superimposed pilot vs. conventional pilots for channel estimation, in *Fortieth Asilomar Conference on Signals Systems and Computers 2006 ACSSC '06*, (Pacific Grove, California USA, 2006), pp. 767–771
4. M Ghogho, DC McLernon, E Alameda-Hernandez, A Swami, Channel estimation and symbol detection for block transmission using data-dependent superimposed training. IEEE Signal Process Lett. 12(3), 226–229 (2005)
5. DC McLernon, E Alameda-Hernandez, AG Orozco-Lugo, MM Lara, Performance of data-dependent superimposed training without cyclic prefix. Electron Lett. 42(10), 604–606 (2006). doi:10.1049/el:20060127
6. E Gayosso-Rios, MM Lara, AG Orozco-Lugo, DC McLernon, Symbol-blanking superimposed training for orthogonal frequency division multiplexing systems, in *7th International Symposium on Wireless Communications Systems (ISWCS)*, York, UK, pp. (19–22 Sept 2006), 204–208
7. C-T Lam, DD Falconer, F Danilo-Lemoine, R Dinis, Channel estimation for SC-FDE systems using frequency domain multiplexed pilots, in *IEEE 64th Vehicular Technology Conf 2006 VTC-2006 Fall*, (Montreal, Canada, 2006), pp. 1–5
8. T Levanen, J Talvitie, M Renfors, Performance evaluation of a DDST based SIMO SC system with PAPR reduction, in *6th International Symposium on Turbo Codes & Iterative Information Processing, ISTC 2010*, (Brest, France, 2010), pp. 186–190
9. DC McLernon, E Alameda-Hernandez, A Orozco-Lugo, MM Lara, New results for channel estimation via superimposed training, in *Proc Second International Symposium on Communications Control and Signal Processing ISCCSP 2006*, (Marrakech, Morocco, 2006). (Article ID cr1001). ISBN: 2-908849-17-8
10. R Raich, H Qian, GT Zhou, Optimization of SNDR for amplitude-limited nonlinearities. IEEE Trans Com-mun. 53(11), 1964–1972 (2005)
11. C Rapp, Effects of HPA-nonlinearity on a 4-DPSK/OFDM-signal for a digital sound broadcasting system, in *Second European Conference on Satellite Communications, ECSC-2*, (Liege, Belgium, 1991), pp. 179–184
12. JJ Bussgang, Crosscorrelation functions of amplitude-distorted gaussian signals, (Technical report (Massachusetts Institute of Technology. Research Laboratory of Electronics), 1952) Report no.: 216,
13. TB Sorensen, PE Mogensen, F Frederiksen, Extension of the ITU channel models for wideband (OFDM) systems, in *IEEE 62nd Vehicular Technology Conference 2005 (VTC-2005-Fall)*, (Dallas, Texas, USA, 2005), pp. 392–396
14. Y Yang, T Ihalainen, M Rinne, M Renfors, Frequency-domain equalization in single-carrier transmission: filter bank approach. EURASIP J Adv Signal Process (2007). **2007**, (Article ID 10438)
15. MV Clark, Adaptive frequency-domain equalization and diversity combining for broadband wireless communications. IEEE J Sel Areas Commun. 16(8), 1385–1395 (1998). doi:10.1109/49.730448
16. E Alameda-Hemndez, DC McLernon, AG Orozco-Lugo, MM Lara, M Ghogho, Improved synchronization for superimposed training based channel estimation, in *IEEE/SP 13th Workshop on Statistical Signal Processing*, (Bordeaux, France, 2005), pp. 1324–1329
17. SMA Moosvi, DC McLernon, AG Orozco-Lugo, MM Lara, M Ghogho, Carrier frequency offset estimation using data-dependent superimposed training. IEEE Commun Lett. 12(3), 179–181 (2008)
18. Y Yang, T Ihalainen, M Renfors, Filter bank based frequency domain equalizer in single carrier modulation, in *Proc 14th IST Mobile & Wireless Communications Summit*, (Dresden, Germany, 2005)
19. M Pukkila, Iterative Receivers and Multichannel Equalisation for Time Division Multiple Access Systems, Ph.D. dissertation, (Helsinki University of Technology, Espoo, Finland, 2003) ISBN 951-22-6717-9
20. R Carrasco-Alvarez, R Parra-Michel, AG Orozco-Lugo, JK Tugnait, Enhanced channel estimation using superimposed training based on universal basis expansion. IEEE Trans Signal Process. 57(3), 1217–1222 (2009)
21. T Levanen, M Renfors, Improved performance bounds for iterative IC LMMSE channel estimator with SI pilots, in *21st Annual IEEE International Symposium on Personal Indoor and Mobile Radio Communications*, (Istanbul, Turkey, 2010), pp. 9–14
22. 3GPP TS36.101 V10.1.0 (2010-12), 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception (Release 10) http://www.3gpp.org/ftp/Specs/archive/36_series/36.101/36101-a10.zip (2012)
23. R Dinis, C-T Lam, D Falconer, Joint frequency-domain equalization and channel estimation using superimposed pilots, in *IEEE Wireless Communications and Networking Conference (WCNC)*, (Las Vegas, NV, USA, 2008), pp. 447–452
24. T Levanen, J Talvitie, M Renfors, Improved performance analysis for super imposed pilot based short channel estimator, in *IEEE International Workshop on Signal Processing Advances for Wireless Communications SPAWC 2010*, (Marrakech, Morocco, 2010), pp. 1–6
25. C Berrou, A Glavieux, P Thitimajshima, Near shannon limit error-correcting coding and decoding: turbo-codes, in, in *IEEE International Conference on Communications*, vol. 2. (Geneva, 1993), pp. 1064–1070
26. P Robertson, E Villebrun, P Hoeher, A comparison of optimal and sub-optimal MAP decoding algorithms operating in the log domain, in *Proc IEEE International Conference on Communications ICC 95 Gateway to Globalization*, vol. 2. (Seattle, WA, USA, 1995), pp. 1009–1013
27. S Sharma, S Attri, RC Chauhan, A simplified and efficient implementation of FPGA-based turbo decoder, in *2003 IEEE International Performance Computing and Communications Conference*, (Longowal, Sangru, 2003), pp. 207–213
28. CL Kei, WH Mow, Improved stopping criteria for iterative decoding of short-frame multi-component turbo codes, in *Proc IEEE International Conference on Communications Circuits and Systems and West Sino Expositions*, vol. 1. (Chengdu, Sichuan, China, 2002), pp. 42–45
29. D Divsalar, F Pollara, Turbo codes for PCS applications, in *Proc IEEE International Conference on Communications ICC'95 Gateway to Globalization*, vol. 1. (Seattle, WA, USA, 1995), pp. 54–59
30. 3rd Generation Partnership Project. 3GPP TS 25.212 V7.2.0 (2006-09); 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Multiplexing and channel coding (FDD) (Release 7) ftp://ftp.3gpp.org/specs/2006-09/Rel-7/25_series/25212-720.zip (2011)
31. R Gold, Optimal binary sequences for spread spectrum multiplexing. IEEE Trans Inf Theory. 13(4), 619–621 (1967)

Publication P5

T. Levanen and M. Renfors, "Improved Energy Efficiency for Wireless SC MIMO Through Data-Dependent Superimposed Training", in *the 5th International Symposium on Communications, Control, and Signal Processing, ISCCSP 2012*, Rome, Italy, 2-4 May 2012.

Copyright ©2012 IEEE. Reprinted, with permission from ISCCSP.

Improved Energy Efficiency for Wireless SC MIMO Through Data-Dependent Superimposed Training

Toni Levanen and Markku Renfors
Department of Communications Engineering
Tampere University of Technology
P.O.Box 553, FIN-33101, Finland
Email: {firstname.lastname}@tut.fi

Abstract—In this paper, we compare the energy efficiency of a single carrier multiple-input multiple-output communications in terms of the energy per data bit required to achieve desired performance level or throughput. The comparison on energy efficiency is done between traditional time-domain multiplexed training and a more recently introduced data-dependent super-imposed training. We extend our earlier single-input multiple-output system to the multiple-input multiple-output case and show how the data-dependent superimposed training based system can achieve better energy efficiency in small and diversity enabled multiple-input multiple-output links. In addition, we present analytical mean squared error results for DDST based MIMO channel estimation and propose methods to improve their accuracy when using short channel estimators.

Keywords: data-dependent superimposed training, channel estimation, energy efficiency, multiple-input multiple-output communications

I. Introduction

Multiple-input multiple-output (MIMO) communications provide the possibility to increase the throughput of a wireless link simply by adding transmit-receive (Tx-Rx) antenna pairs. The capacity of a wireless MIMO link increases linearly with the number Tx-Rx antenna pairs (assuming independently fading channels) [1]. The main problem in achieving this capacity with traditional time-domain multiplexed training (TDMT) is the fact that we have to allocate more time slots for the training sequence to enable the estimation of all wireless channels involved in the MIMO communication link. Thus, as each Tx-Rx antenna pair increases capacity, it also increases the amount of training information required to estimate the channel between the new Tx antenna and all of the Rx antennas.

Another approach for channel estimation, that has recently obtained growing interest, is the data-dependent superimposed training (DDST) based channel estimation (see, for example [2] and [3]). In DDST, all the time slots are dedicated for the user data symbols and the training sequence is arithmetically added on top of the user data symbols. The data-dependent portion of the training sequence is typically related to the cyclic mean of the transmitted user data symbols. The removal

of the cyclic mean corresponds to removing certain frequency bins from the discrete Fourier-transform (DFT) of the user data symbol sequence. Then, the cyclic pilot sequence is contained in these frequency bins and detected in the receiver without interference from the user data. In the receiver, the missing information can be effectively estimated by iteratively calculating the cyclic mean of the detected symbols [4]. For an interested reader, superimposed pilot based single carrier multiuser/MIMO communications, without the data-dependent component, is studied, for example, in [5] and [6].

In this paper, we compare the energy consumption in terms of average energy per bit over the one sided noise spectral density (E_b/N_0). We show that with DDST, in single carrier multiantenna transmission, one can further improve the energy efficiency in single-input multiple-output (SIMO) scenarios and in MIMO communications. As we increase the number of Tx-Rx pairs or the size of the constellation, the performance gain of the DDST decreases with respect to the TDMT, because the effective SNR of a data symbol with DDST decreases as larger portion of average power is removed (due to increased length of pilot sequence) or the deviation of the cyclic mean component increases with respect to the symbol distances (due to increased size of the constellation).

This paper is organized as follows. In Section II the system model and receiver architecture are introduced as extension of the SIMO model of [4] without any power constraints. The channel estimation and related analysis is presented in Section III. In Section IV we compare the energy efficiency of DDST and TDMT with different antenna configurations. Furthermore, we comment on the performance limiting factors of DDST and try to give an idea of the parameter set in which DDST can overperform TDMT. Finally, in Section V, our conclusion are presented.

II. System Model

The system model originates from a possible future uplink-direction mobile wireless link. Therefore, most of the signal processing is located in the receiver side (in the base station). The conceptual block diagram of the simulated system model is given in Fig. 1.

In the transmitter side, we have traditional signal processing blocks, as channel encoder, channel interleaver, symbol mapper and transmit pulse shape filtering. In the studies considered
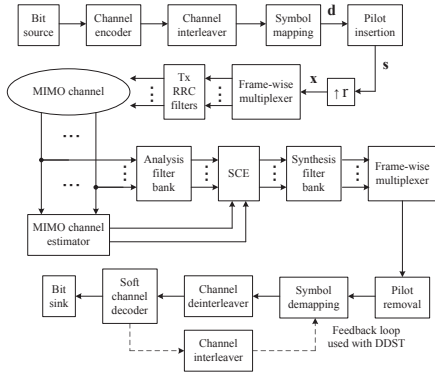
Fig. 1. System model for the single carrier based multiple-input multiple-output simulations.

in this paper, we have concentrated on spatial multiplexing and have not considered any space-time codes. In the receiver side, the maximum number of antennas is limited to four. The MIMO channel frequency domain equalization is done with a filter bank based linear equalizer with MSE criteria, as presented in [7].

The power of the MIMO signal is normalized to unity, thus each Tx antenna transmits with average power equal to $\sigma_{s,i}^2 = 1/N_{Tx}$, where i is the index for the transmit antenna. The received signal-to-noise ratio (SNR), based on the given energy per bit over the one sided noise spectral density (E_b/N_0), is given as

$$SNR = \frac{N_b}{r N_s} \frac{E_b}{N_0}, \tag{1}$$

where N_b is the number of user data bits per frame, N_s is the number of symbols per frame and r is the oversampling factor used in the receiver. Note that in every case the number of user data bits is bigger with DDST than with TDMT, and the difference increases as we increase the number of Tx-Rx pairs. This is because we allocate more symbols for training in TDMT that leads to less user data symbols per frame.

Given a certain user data symbol sequence $\mathbf{d}_i$, transmitted from i^{th} Tx antenna, taken from some constellation $\mathbf{d}_i \in \mathcal{C}^N$, where N is the length of the sequence, we can define the symbols after inserting pilots as $\mathbf{s}_i = [\mathbf{p}_i^T \mathbf{d}_i^T]^T$ or $\mathbf{s}_i = \sqrt{(1-\gamma)/(1-1/N_c)}\mathbf{J}\mathbf{d}_i + \sqrt{\gamma}\mathbf{p}_{c,i}$, for TDMT or DDST, respectively. We assume that both, the data sequence $\mathbf{d}_i$ and the cyclic pilot sequence $\mathbf{p}_{c,i}$, have a unit variance $\sigma_d^2 = 1$ and $\sigma_{p_{c,i}}^2 = 1$, respectively. With DDST, γ defines the fraction of total power allocated to the cyclic pilot sequence $\mathbf{p}_{c,i} = \mathbf{1}_{N_c \times 1} \otimes \mathbf{p}_i$, where $\otimes$ is the Kronecker product and $\mathbf{p}_i$ is a certain cyclic shift of the known basis pilot sequence $\mathbf{p}$, transmitted from the ith Tx antenna. The cyclic mean of the data sequence is removed by multiplying with matrix $\mathbf{J} = \mathbf{I} - 1/N_c \mathbf{1}_{N_c} \otimes \mathbf{I}_{N_p}$. In this paper we assume that the frame length is $N = N_c N_p$, where N_c is the number of cyclic copies and N_p is the length of the basis pilot sequence $\mathbf{p}$. Note, that

the basis pilot sequence length is now N_{Tx} times the channel estimator length, allowing us to estimate the channels between all Tx antennas and a single Rx antenna. The signal transmitted from Tx antenna i with DDST is defined as

$$\begin{aligned}\mathbf{x}_i \quad &= \mathbf{H}_{RRC}\sqrt{1/N_{Tx}} \\ &(\sqrt{(1-\gamma)/(1-1/N_c)}\mathbf{J}\mathbf{d}_i + \sqrt{\gamma}\mathbf{p}_{c,i}) \otimes [r \underbrace{0\ldots0}_{r-1 \text{ zeros}}],\end{aligned} \tag{2}$$

where $\mathbf{H}_{RRC}$ is a circular matrix containing the Tx pulse shape filter coefficients and is used to calculate the convolution between the Tx pulse shape filter and the symbol sequence. The transmitted symbol sequence is normalized with a factor $1/(1-1/Nc)$, to take into account the average power loss caused by the removal of the cyclic mean component. Kronecker product with vector $\mathbf{r} = [r\ 0\ldots0]$ realizes the r times oversampling with power normalization.

The used basis pilot sequences, $\mathbf{p}$, are so called chirp sequences [8], which are shown to be optimal for DDST in [9] (noted as optimal channel independent (OCI) pilot sequences) and also for TDMT in MIMO communications [10]. The same training is used with both, TDMT and DDST. The only difference is in the ordering of the pilots to each transmitted frame, in order to obtain a full pilot matrix in the receiver for MIMO channel estimation.

In the receiver side, we have channel estimation, filter bank based frequency domain equalization implementing a close-to-optimal linear MIMO detector with heavily frequency-selective channels, pilot removal, symbol demapping and decoding. With DDST, we have a soft iterative loop that is used to estimate the removed cyclic mean of the user data sequence, similar to what was proposed in [4]. The signal received in Rx antenna with index k is defined as

$$\mathbf{y}_k = [\mathbf{H}_{k,1}\ \mathbf{H}_{k,2}\ldots\mathbf{H}_{k,N_{Tx}}][\mathbf{x}_1^T\ \mathbf{x}_2^T\ldots\mathbf{x}_{N_{Tx}}^T]^T + \mathbf{n}_k, \tag{3}$$

where $\mathbf{H}_{k,i}$ represents the convolution matrix of the channel response between i^{th} Tx antenna and k^{th} Rx antenna and $\mathbf{n}_k$ is a vector of complex Gaussian noise components with variance $\sigma_n^2 = 1/SNR$. We have normalized the power response of all channel realization to unity.

We have assumed two times oversampling in the receiver $r = 2$, which allows us to efficiently incorporate the RRC filtering in the subchannel wise equalization (SCE) used in the filter bank. The different spatial data streams are obtained by applying the linear subcarrier wise MIMO equalizer to the received signal [7]. The number of subbands in the analysis filter bank is set to 1024 and in the synthesis filter bank it is then 512, due to $r = 2$ times down sampling. The used overlapping factor in the filter bank is equal to 5 and the filter bank roll-off is equal to 1. More details on the filter bank can be found, e.g., in [7], [11] and references therein.

After the channel equalization, we remove the training symbols from the received sequence and normalize the average power per layer to $\sigma_{\hat{d}}^2 = 1 + \sigma_n^2$ with TDMT and

to $\sigma_d^2 = (1 - 1/N_c) + \sigma_n^2$ with DDST. We have assumed that the noise variance σ_n^2 is known by the receiver. With DDST, after the normalization, we add the first cyclic mean estimate to the received symbol sequence. This cyclic mean estimate is based on the hard symbol estimates before channel decoder, as presented in [2]. Then we proceed to the soft channel decoder and with DDST we use maximum of 3, 8, or 10 feedback iterations for soft cyclic mean estimation with 4-, 16-, or 64-QAM modulation, respectively. The soft cyclic mean estimation at each feedback iteration is based on the soft symbol estimates generated from the soft coded bit estimates from the turbo decoder. More details on the soft symbol mapping and demapping can be found, e.g., in [12].

Finally, after all the iterations, we provide the detected bits to the bit sink, from which we eventually obtain bit error rate (BER) and block error rate (BLER) results. We have used BLER to represent goodness of performance of the desired training method and not the frame error rate, because we have defined that each transmitted frame contains q coded blocks, where q is equal to the number of bits per symbol. Thus, each binary coded block in a frame with any constellation is of the same size and corresponds to the number of data symbols per frame. This way the block wise decoding complexity is constant with each constellation and the decoding process can be parallelized in the receiver with larger constellations.

III. CHANNEL ESTIMATION AND MSE ANALYSIS

In this Section, we introduce the used channel estimator model and derive the related analytic channel estimation MSE. We have used a LS-LMMSE channel estimator, which first uses a least-squares (LS) channel estimator to obtain initial channel estimates and then uses a linear minimum mean-squared-error (LMMSE) channel estimator, based on the LS estimate, to obtain improved channel estimates. The LS channel estimate, for all channels between all Tx antennas and k^{th} Rx antenna is defined as

$$\hat{\mathbf{h}}_{LS,k} = \frac{\mathbf{P}_r^H \sqrt{N_{Tx}}}{r^2 N_p \sigma_p^2} \hat{\mathbf{m}}_{y,k} = \mathbf{h}_k + \frac{\mathbf{P}_r^H \mathbf{H}_{RRC} \hat{\mathbf{m}}_{n,k}}{r^2 N_p \sigma_p^2}, \quad (4)$$

where $\mathbf{h}_k = [\mathbf{h}_{k,1}^T \ \mathbf{h}_{k,2}^T \ldots \mathbf{h}_{k,N_{Tx}}^T]^T$ is the equivalent channel between all Tx antennas and k^{th} Rx antenna, $\mathbf{P}_r^H$ is a cyclic pilot matrix oversampled with factor $r = 2$ and having basis pilot vector $\mathbf{p}$ as its first column vector, $\hat{\mathbf{m}}_{y,k} = \mathbf{J}_{Rx} \mathbf{y}_k$ is the cyclic mean of the received signal vector in antenna k, and the matrix used to calculate the cyclic mean of the received sequences is defined as $\mathbf{J}_{Rx} = 1/N_c \mathbf{1}_{1 \times (N_c+1)} \otimes \mathbf{I}_{N_p}$. This matrix is an extended version of the one used in the transmitter, because the received signal is longer in time because of the time dispersive channel. Similarly, $\hat{\mathbf{m}}_{n,k} = \mathbf{J}_{Rx} \mathbf{n}_k$, is the cyclic mean of the received noise component of antenna k. Note, that vector $\hat{\mathbf{h}}_{LS,k} = [\hat{\mathbf{h}}_{LS,k,1}^T \ \hat{\mathbf{h}}_{LS,k,2}^T \ldots \hat{\mathbf{h}}_{LS,k,N_{Tx}}^T]^T$ contains now the channel estimates from all the Tx antennas to the kth Rx antenna. See [10] for further details how the channel estimator is able to obtain channel estimates based on multiple-input single-output (MISO) principle with TDMT.

The same ideology holds for DDST and has been used for the results presented in this paper.

In the channel estimator, we approximate the diagonal correlation matrix $\mathbf{C}$ by the instantaneous tap power obtained from the LS channel estimator as

$$\mathbf{C}_{\hat{\mathbf{h}}_{LS,k}} = diag\left\{ |\hat{h}_{LS,k}(0)|^2, |\hat{h}_{LS,k}(1)|^2, \cdots, \right. \\ \left. |\hat{h}_{LS,k}(rN_p - 1)|^2 \right\}. \quad (5)$$

By assuming the cyclic chirp (OCI) training sequence, the LS-LMMSE estimator can be reduced to

$$\hat{\mathbf{h}}_{LS-LMMSE,k} = \frac{\mathbf{P}_r^H \sqrt{N_{Tx}}}{(\sigma^2 \mathbf{C}_{\hat{\mathbf{h}}_{LS}}^{-1} + r^2 N_p \sigma_p^2 \mathbf{I}_{rN_p \times rN_p})} \hat{\mathbf{m}}_{y,k}. \quad (6)$$

The variable σ^2 corresponds to the total interference power on top of the cyclic mean of the received signal and is given as

$$\sigma^2 = N_{Tx}(1/N_c + 1/N_c^2)\sigma_w^2 \|\mathbf{h}_{RRC}\|^2. \quad (7)$$

Let us define an error vector $\mathbf{e}_{LS,k} = \hat{\mathbf{h}}_{LS,k} - \mathbf{h}_k(\mathbf{\Xi})$, representing the channel estimation error between the estimated channel and a fraction of the true channel corresponding to the estimated part. Because we have not estimated the full equivalent channels, we use indexing set $\mathbf{\Xi}$ to define the samples that are included in the channel estimation process. This idea of short channel estimator was presented in [13], where it was studied with superimposed training. More discussion on the numerical values used is given in Section IV. Now, assuming that the channel taps are i.i.d., the channel estimation mean-squared-error (MSE) for DDST with LS channel estimator is defined as

$$MSE_{LS,k} = trE[\mathbf{e}_{LS,k}\mathbf{e}_{LS,k}^H] = \ldots = \frac{(1/N_c + 1/N_c^2)\sigma_n^2 N_{Tx}}{\gamma r}, \quad (8)$$

and the average MSE over all spatial channels is defined as

$$MSE_{LS} = \frac{1}{N_{Tx}N_{Rx}} \sum_{k=1}^{N_{Rx}} MSE_{LS,k}. \quad (9)$$

In similar manner, for DDST with the LS-LMMSE channel estimator, we can define the channel estimation MSE between all Tx antennas and k^{th} Rx antenna as

$$MSE_{LS-LMMSE,k} \\ = trE[\mathbf{e}_{LS-LMMSE,k}\mathbf{e}_{LS-LMMSE,k}^H] = \ldots = \\ \sum_{l=1}^{N_p} \frac{\sigma_n^4 \sigma_{LS,k}^{-4}(l)}{(\sigma_n^2 \sigma_{LS,k}^{-2}(l) + \gamma r^2 N_p)^2} \sigma_{h,k}^2(l) + \frac{\gamma r^2 N_p N_{Tx}(1 + 1/N_c)\sigma_n^2}{(\sigma_n^2 \sigma_{LS,k}^{-2}(l) + \gamma r^2 N_p)^2}, \quad (10)$$

,where $\sigma_{h,k}^2(l) = E[|h_k(l)|^2]$ is the expected power of the lth tap in equivalent channel $\mathbf{h}_k$ and $\sigma_{LS,k}^2 = \sigma_{h,k}^2(l) + MSE_{LS,k}/N_p$ is the expected power of the lth tap from the LS
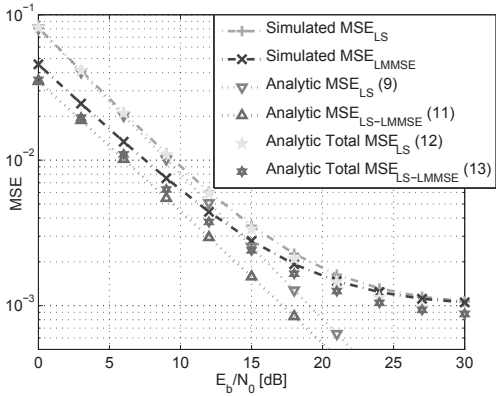
Fig. 2. Channel estimation MSE comparison, for DDST with 2 Tx antennas and 4 Rx antennas, between the simulated MSE and the analytical MSE obtained from (9), (11), (12), and (13).

channel estimator. The average MSE over all spatial channels is defined as

$$MSE_{LS-LMMSE} = \frac{1}{N_{Tx}N_{Rx}} \sum_{k=1}^{N_{Rx}} MSE_{LS-LMMSE,k}. \quad (11)$$

In Fig. 2, we compare the simulated channel estimation MSE with analytical results based on (9) and (11). In this case we have used 2 Tx antennas and 4 Rx antennas. As we can see, there is a error floor in the simulated values which is not present in the analytical ones. This same phenomenon was noted also in [13], which it was proposed that the total MSE of the short channel estimator should include also the expected power of the non-estimated channel taps. Let us now define the analytic MSE estimates based on this idea as

$$Total\ MSE_{LS} = MSE_{LS} + \sum_{s \in \Psi/\Xi} |h_k(s)|^2 \text{and} \quad (12)$$

$$Total\ MSE_{LS-LMMSE} = MSE_{LS-LMMSE} + \sum_{s \in \Psi/\Xi} |h_k(s)|^2, \quad (13)$$

where $\Psi = [0, 1, \ldots, L-1]$ is the indexing set containing all possible index values of the equivalent channel response $\mathbf{h}_k$. The numerical values for our simulations are defined in Section IV.

We have plotted also these MSE results in Fig. 2, and have named them as Total MSE_{LS} and Total $MSE_{LS-LMMSE}$, for the LS and LS-LMMSE channel estimators, respectively. In [13], the intuition behind the idea is that the channel taps not estimated by the short channel estimator fold on the estimated portion in the cyclic mean computation, and therefore cause the error floor. In our simulations, we have noticed the same phenomenon also with short channel estimation with TDMT, in which the MSE error floor corresponds quite accurately to the expected power of the non-estimated channel taps.

IV. PERFORMANCE COMPARISON BETWEEN DDST AND TDMT

In this section we compare the performance of the SC based MIMO communications with DDST and TDMT in terms of the required average E_b/N_0 to achieve BLER = 10^{-2} or a predefined rate.

The used channel model is the block fading extended ITU Ped. B channel profile [14], which is of length 115 samples with symbol rate $f_{symbol} = 15.36$ MHz and sample rate $f_{sample} = 2f_{symbol}$. In addition, the equivalent channel includes twice the root-raised-cosine (RRC) pulse shape filtering of order $N_{RRC} = 64$, making the equivalent channel length per spatial channel 243 samples long, and the overall channel response observed in the kth receive antenna is of length $L = 243N_{Tx}$ samples. The used RRC filter has a roll-off factor equal to 0.1.

With DDST, we have decided to use channel estimators of length $N_p = (60N_{Tx})$ symbols, if $N_{Tx} = 1, 2$, or 4, and $N_p = (64N_Tx)$ symbols, if $N_{Tx} = 3$. This is to ensure that $N_c = N/N_p$ is an integer. With TDMT the length of the pilot sequence is $N_p = (N_{Tx} + 1) * 60 - 1$. This means that we estimate only 120 or 128 samples per spatial channel. This kind of short channel estimator allows us to increase the maximum throughput of the TDMT and to increase the number of cyclic copies with DDST. The channel estimation degradation due to this choice is negligible. As an example, the indexing set for samples estimated with 1,2 or 4 Tx antennas, is then given as

$$\Xi = \underbrace{[1\ 1\ldots 1]}_{N_{Tx}\ \text{ones}} \otimes [N_{RRC}/2\ N_{RRC}/2 + 1\ldots$$
$$N_{RRC}/2 + 119] + 243[0\ 1\ldots N_{Tx} - 1] \otimes \mathbf{1}_{1\times 120}. \quad (14)$$

In other words, in the channel estimation process, we ignore the first $N_{RRC}/2$ samples and all the samples after index $N_{RRC}/2 + 119$ for each spatial channel.

The used codec in the simulations is a turbo code [15] with generator matrix $G = [1\ \frac{1\ 5}{1\ 3}]$ and the decoding algorithm is the max-log-MAP algorithm with extrinsic information weighting by a factor $\mu = 0.75$ [16]. We have allowed maximum of 5 decoding iterations per code block in the turbo decoder. The channel interleaver and the interleavers inside the turbo codec are S-interleavers [17].

The performance of the DDST based transmission depends on the constellation, number of the cyclic copies of the pilot sequence and on the code rate R. We have done simulations with 4-, 16-, and 64-QAM constellations and with code rates $R = 0.5$, $R = 0.67$, and $R = 0.75$. Let us now discuss on how the different simulation parameters effect on the DDST performance.

As we increase the size of the constellation, it becomes more sensitive to the distortion caused by the removal of the cyclic mean and to the channel estimation errors. Thus, with bigger constellation, higher number of cyclic copies is required to achieve desired level of performance. Increasing pilot power
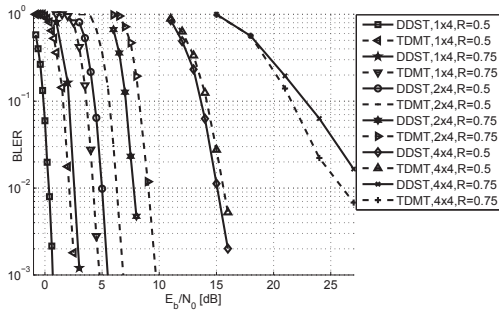
Fig. 3. Block error rate performance with 16-QAM constellation with different MIMO antenna configurations and code rates $R = 0.5$ and $R = 0.75$.
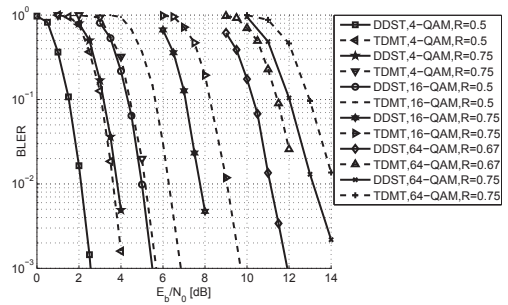

Fig. 4. Block error rate performance with 2x4 MIMO antenna configuration with 4-, 16-, and 64-QAM modulations with code rates $R = 0.5$, $R = 0.67$ and $R = 0.75$.

in this case is a two edged blade, because as we improve the channel estimation MSE, we decrease the effective signal-to-interference-and-noise ratio in the equalizer output. In our simulations, we noticed that the value $\gamma = 0.1$ works rather well in all of the cases. As a rule of thumb, one can consider of allocating the same average power for DDST pilots as is allocated for TDMT pilots, because it was shown in [2], that this leads to the same channel estimation MSE with least-squares type of channel estimators.

The number of cyclic copies depends on the equivalent channel length, channel estimator length and frame length. If the equivalent channel is short with respect to the transmitted frame, we can have higher number of cyclic copies per frame, and thus improve the performance with DDST through improved noise averaging in the channel estimation process and decreased variance of the interference term caused by the cyclic mean removal. If the expected equivalent channel has most of its power concentrated in a relatively short time interval, we can use a short channel estimator that estimates only the significant portion of the channel. With extended ITU Ped. B channel model, we can collect 99.86% of the total power per layer with the used channel estimator of length 120 samples.

Finally, the code rate R affects the performance of the iterative cyclic mean estimation process. The iterative, soft feedback scheme is effective especially with larger constellations and with lower code rates. This is because larger constellations are more sensitive to the cyclic mean estimation errors and smaller code rates allows the turbo decoder to provide more new information for the cyclic mean estimation.

With these short explanations on the expected performance of DDST, we first look at an example of the throughput performance with 16-QAM modulation in different antenna configurations, shown in Fig. 3. In the presented cases the throughput performance of a DDST based system exceeds that of a TDMT based system when we have receiver diversity. In the 4×4 MIMO case DDST is marginally better with coding rate $R = 0.5$, but TDMT provides better performance with $R = 0.75$. One can notice that the difference between DDST

and TDMT based systems decreases as we increase the number of Tx-Rx pairs and code rate, and how TDMT eventually exceeds DDST in the 4×4 MIMO case with $R = 0.75$. In 6x6 or 8x8 MIMO antenna configurations TDMT would provide better energy efficiency than DDST, while keeping the other parameters fixed.

In Fig. 4, we compare the throughput performance with different training schemes in 2x4 MIMO antenna configuration. As expected, we can see how the TDMT system starts to achieve the performance of DDST as the constellation size increases or with higher code rates. From the maximal throughput point of view, increasing the number of Tx-Rx pairs and using lower modulation is a more energy efficient and robust method than using constellations larger or equal to 64-QAM.

In Table IV, we have given the best constellation, antenna configuration and code rate combination to achieve the given throughput and the required E_b/N_0 for both, DDST and TDMT. Note that we have simulated discontinuous block fading channel, so these results should be considered as a rough estimate of the required E_b/N_0 to achieve certain throughput in a continuous transmission channel. In all of the cases, the two systems achieve the best E_b/N_0 with the same constellation, code rate and antenna configuration set. We can notice how the TDMT achieves similar E_b/N_0 requirement with high throughput requirement (≥ 150 Mb/s). If we would have more Rx antennas, the DDST based system would be better at even higher throughput rates, given that we could use smaller constellations or provide reception diversity with larger ones. In [6], an iterative receiver for superimposed training was studied, and its performance was compared to coordinated and uncoordinated TDMT users. Also, in [6], the performance of the superimposed training based system was able to compete with the coordinated TDMT based system, if there was reception diversity available. This results is comparable with ours, because we have assumed perfect synchronization for both, DDST and TDMT users. Compared to the superimposed training considered in [6], DDST based system is more sensitive to the frame synchronization, as was shown in [3]. On the other hand, DDST provides better initial

TABLE I
BEST SETUP AND REQUIRED E_b/N_0 FOR DESIRED THROUGHPUT

Throughput	DDST	TDMT
>50 Mb/s	16-QAM, 2x4, R=0.5, E_b/N_0= 4.5 dB	16-QAM, 2x4, R=0.5, E_b/N_0= 6 dB
>100 Mb/s	16-QAM, 3x4, R=0.67, E_b/N_0= 10.5 dB	16-QAM, 3x4, R=0.67, E_b/N_0= 11.5 dB
>150 Mb/s	64-QAM, 3x4, R=0.67, E_b/N_0=18	64-QAM, 3x4, R=0.67, E_b/N_0=18

channel estimates, because the user data does not affect the channel estimation MSE.

V. CONCLUSION

In this paper, we have proposed an analytic channel estimation MSE for DDST in MIMO communications for the LS and LS-LMMSE channel estimator and compared its performance to the simulated channel estimation MSE. In addition, the error floor noted in the simulated MSE results seems to correspond to the sum power of the expected channel taps outside the short channel estimator.

Furthermore, we have compared the required E_b/N_0 with either DDST or TDMT based channel estimation in SC MIMO communications. We have shown, that DDST can overperform the traditional TDMT in small MIMO scenarios (less than 4 Tx-Rx pairs) and when sufficient receiver diversity is available. As typical, there is a complexity penalty to pay for the improved efficiency. Fortunately, the power consumption of the added complexity in the receiver hardware will become smaller than the savings achieved in the transmission power, as the hardware evolves following the famous Moore's law.

The scenario, in which DDST overperforms TDMT, depends on the number of antennas, frame length and on the channel estimator length (or channel length). As long as we have sufficient number of cyclic copies of the basis pilot sequence, we do not cause significant interference on the data symbols and we obtain improved channel estimates through the noise averaging in the DDST based channel estimation. Thus, even though the maximum throughput of the TDMT is significantly smaller in larger MIMO communications (e.g., $k \times k$ MIMO cases, where $k \geq 4$), the interference caused by removing the cyclic mean with DDST may cancel the possible throughput benefits, making TDMT a more viable solution in these cases.

REFERENCES

[1] G. J. Foschini, "Layered Space-Time Architecture for Wireless Communication in a Fading Environment when Using Multi-Element Antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.

[2] M. Ghogho et al., "Channel Estimation and Symbol Detection for Block Transmission Using Data-Dependent Superimposed Training," *IEEE Signal Process. Lett.*, vol. 12, no. 3, pp. 226–229, 2005.

[3] E. Alameda-Hernandez et al., "Frame/Training Sequence Synchronization and DC-Offset Removal for (Data-Dependent) Superimposed Training Based Channel Estimation," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2557–2569, 2007.

[4] T. Levanen, J. Talvitie, and M. Renfors, "Performance Evaluation of a DDST Based SIMO SC System With PAPR Reduction," in *6th International Symposium on Turbo Codes & Iterative Information Processing, ISTC 2010,*, Brest, France, 6-10 Sept.

[5] S. He, J. K. Tugnait, and X. Meng, "On Superimposed Training for MIMO Channel Estimation and Symbol Detection," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 3007–3021, 2007.

[6] J. K. Tugnait and S. He, "Multiuser/MIMO Doubly Selective Fading Channel Estimation Using Superimposed Training and Slepian Sequences," *IEEE Trans. Veh. Technol.*, vol. 59, no. 3, pp. 1341–1354, 2010.

[7] T. Ihalainen et al., "Channel Equalization for Multi-Antenna FBMC/OQAM Receivers," *IEEE Trans. Veh. Technol.*, vol. 60, no. 5, pp. 2070–2085, June 2011.

[8] D. Chu, "Polyphase Codes with Good Periodic Correlation Properties," *IEEE Trans. Inf. Theory*, vol. 18, no. 4, pp. 531–532, Jul. 1072.

[9] A. G. Orozco-Lugo, M. M. Lara, and D. C. McLernon, "Channel Estimation Using Implicit Training," *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 240–254, 2004.

[10] O. Weikert and U. Zölzer, "Efficient MIMO Channel Estimation With Optimal Tranining Sequences," in *in Proc. 1st Workshop on Commercial MIMO-Components and -Systems (CMCS 2007)*, Duisburg, Germany, 13-14 Sept. 2007.

[11] Y. Yang, T. Ihalainen, M. Rinne and M. Renfors, "Frequency-domain Equalization in Single-carrier Transmission: Filter Bank Approach," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, Article ID 10438, 16 pages, 2007.

[12] C. Laot, A. Glavieux, and J. Labat, "Turbo Equalization: Adaptive Equalization and Channel Decoding Jointly Optimized," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 9, pp. 1744–1752, 2001.

[13] T. Levanen, J. Talvitie and M. Renfors, "Improved Performance Analysis for Super Imposed Pilot Based Short Channel Estimator," in *IEEE International Workshop on Signal Processing Advances for Wireless Communications, SPAWC 2010*, Marrakech, Morocco, 20-23 June 2010.

[14] T. B. Sorensen, P. E. Mogensen, and F. Frederiksen, "Extension of the ITU Channel Models for Wideband (OFDM) Systems," in *62nd IEEE Vehicular Technology Conference, VTC 2005 Fall*, vol. 1, 25-28 Sept. 2005, pp. 392–396.

[15] C. Berrou, A. Glavieux and P. Thitimajshima, "Near Shannon Limit Error-Correcting Coding and Decoding: Turbo-Codes," in *IEEE International Conference on Communications*, vol. 2, Geneva, May 1993, pp. 1064–1070, conference Record.

[16] S. Sharma, S. Attri, and R.C. Chauhan, "A Simplified and Efficient Implementation of FPGA-Based Turbo Decoder," in *2003 IEEE International Performance, Computing, and Communications Conference*, 9-11 Apr. 2003, pp. 207–213.

[17] D. Divsalar and F. Pollara, "Turbo Codes for PCS Applications," in *IEEE International Conference on Communications, ICC'95*, Seattle, WA, USA, 18-22 June 1995, pp. 54–59.