

Linköping Studies in Science and Technology
Dissertations, No. 1756

Phase Noise and Wideband Transmission in Massive MIMO

Antonios Pitarokoilis



Division of Communication Systems
Department of Electrical Engineering (ISY)
Linköping University, SE-581 83 Linköping, Sweden
www.commsys.isy.liu.se

Linköping 2016

Phase Noise and Wideband Transmission in Massive MIMO

© 2016 Antonios Pitarokoilis, unless otherwise noted.

ISBN 978-91-7685-791-5

ISSN 0345-7524

Printed in Sweden by LiU-Tryck, Linköping 2016

To my parents and Kallia

Οὐκ ἴσμεν δὲ τὸ ἀληθὲς ἄνευ τῆς αἰτίας.

Ἀριστοτέλους, Μετὰ τὰ φυσικά

We do not know a truth without its cause.

Aristotle, Metaphysics

Abstract

In the last decades the world has experienced a massive growth in the demand for wireless services. The recent popularity of hand-held devices with data exchange capabilities over wireless networks, such as smartphones and tablets, increased the wireless data traffic even further. This trend is not expected to cease in the foreseeable future. In fact, it is expected to accelerate as everyday apparatus unrelated with data communications, such as vehicles or household devices, are foreseen to be equipped with wireless communication capabilities.

Further, the next generation wireless networks should be designed such that they have increased spectral and energy efficiency, provide uniformly good service to all of the accommodated users and handle many more devices simultaneously. Massive multiple-input multiple-output (*Massive MIMO*) systems, also termed as large-scale MIMO, very large MIMO or full-dimension MIMO, have recently been proposed as a candidate technology for next generation wireless networks. In Massive MIMO, base stations (BSs) with a large number of antenna elements serve simultaneously only a few tens of single antenna, non-cooperative users. As the number of BS antennas grow large, the normalized channel vectors to the users become pairwise asymptotically orthogonal and, therefore, simple linear processing techniques are optimal. This is substantially different from the current design of contemporary cellular systems, where BSs are equipped with a few antennas and the optimal processing is complex. Consequently, the need for redesign of the communication protocols is apparent.

The deployment of Massive MIMO requires the use of many inexpensive and, potentially, off-the-shelf hardware components. Such components are likely to be of low quality and to introduce distortions to the information signal. Hence, Massive MIMO must be robust against the distortions introduced by the hardware impairments. Among the most important hardware impairments is phase noise, which is introduced by local oscillators (LOs)

at the BS and the user terminals. Phase noise is a phenomenon of particular importance since it acts multiplicatively on the desired signal and rotates it by some random and unknown argument. Further, the promised gains of Massive MIMO can be reaped by coherent combination of estimated channel impulse responses at the BS antennas. Phase noise degrades the quality of the estimated channel impulse responses and impedes the coherent combination of the received waveforms.

In this dissertation, wideband transmission schemes and the effect of phase noise on Massive MIMO are studied. First, the use of a low-complexity single-carrier precoding scheme for the broadcast channel is investigated when the number of BS antennas is much larger than the number of served users. A rigorous, closed-form lower bound on the achievable sum-rate is derived and a scaling law on the potential radiated energy savings is stated. Further, the performance of the proposed scheme is compared with a sum-capacity upper bound and with a bound on the performance of the contemporary multi-carrier orthogonal frequency division multiplexing (OFDM) transmission.

Second, the effect of phase noise on the achievable rate performance of a wideband Massive MIMO uplink with time-reversal maximum ratio combining (TR-MRC) receive processing is investigated. A rigorous lower bound on the achievable sum-rate is derived and a scaling law on the radiated energy efficiency is established. Two distinct LO configurations at the BS, i.e., the common LO (*synchronous*) operation and the independent LO (*non-synchronous*) operation, are analyzed and compared. It is concluded that the non-synchronous operation is preferable due to an averaging of the independent phase noise sources. Further, a progressive degradation of the achievable rate due to phase noise is observed. A similar study is extended to a flat fading uplink with zero-forcing (ZF) receiver at the BS.

The fundamental limits of data detection in a phase-noise-impaired uplink are also studied, when the channel impulse responses are estimated via uplink training. The corresponding maximum likelihood (ML) detector is provided for the synchronous and non-synchronous operations and for a general parameterization of the phase noise statistics. The symbol error rate (SER) performance at the high signal-to-noise ratio (SNR) of the detectors is studied. Finally, rigorous lower bounds on the achievable rate of a Massive MIMO-OFDM uplink are derived and scaling laws on the radiated energy efficiency are stated.

Populärvetenskaplig sammanfattning

Det har skett en massiv tillväxt i efterfrågan på trådlösa tjänster de senaste årtiondena. Dessutom blir handhållna enheter med förmåga att kommunicera över trådlösa nätverk, som mobiltelefoner och surfplattor, allt vanligare vilket gör den trådlösa trafiken ökar ytterligare. Denna trend väntas inte upphöra inom överskådlig framtid. I själva verket förväntas den accelerera i takt med att vardagliga apparater tidigare orelaterade till kommunikation, såsom fordon eller hushållsapparater, planeras vara utrustade med trådlösa kommunikationsmöjligheter. Stora leverantörer inom den trådlösa kommunikationsindustrin förutspår en tiofaldig ökning av data-trafiken fram till år 2019 jämfört med 2014 och en tusenfaldig ökning är bara en tidsfråga.

Vidare bör nästa generation trådlösa nätverk utformas så att de har ökad effektivitet i förhållande till den tillhandahållna datahastigheten och energikonsumtionen, jämn kvalitet på tjänster till användarna samt hantera många fler enheter samtidigt. Massiv MIMO ("multiple-input multiple-output"), även kallad hyper MIMO och storskalig MIMO, har nyligen föreslagits som en kandidaterande teknik för nästa generations trådlösa nätverk. I massiv MIMO är basstationen utrustad med ett stort antal antal antenner och betjänar ett tiotal icke samarbetande, användare samtidigt. Med ökande antal antenner vid basstationen minskar störningar mellan kanaler till olika användare och enkla signalbehandlingstekniker har mycket bra prestanda.

Utbyggnaden av massiv MIMO kräver användning av många billiga och potentiellt sett icke specialgjorda komponenter. Sådana komponenter kommer sannolikt att vara av låg kvalitet och införa förvrängningar av informationssignalen. Därför måste Massiva MIMO vara robust mot dessa förvrängningar som kommer från användning av icke-ideal hårdvara.

En av de viktigaste funktionsnedsättningarna från icke-ideal hårdvara är fasbrus. Denna funktionsnedsättning introduceras av komponenter som kallas lokala oscillatorer som finns i basstationen såväl som i användarterminalerna. Fasbrus är ett fenomen av särskild betydelse, eftersom den slumpvis roterar den önskade signalen och kan inte åtgärdas genom att öka effekten för den önskade signalen. Vidare kan de utlovade vinsterna av massiv MIMO skördas genom lämplig kombination av de mottagna signalerna vid basstationen. Emellertid förvränger rotationen från fasbruset denna kombinerings.

I denna avhandling studerar vi bredbandsöverföringssystem och effekten av fasbruset på Massive MIMO. I bredbandsöverföring skiljer sig svaret från kanalen från en signal för varje frekvenskomponent. Detta resulterar i en förvrängning av den mottagna signalen och ett behov för en mottagarstruktur som har förmågan att återvinna den sända signalen med så lite förvrängning som möjligt. I dagens system delas därför bredbandskanalen upp i flera icke-störande smalbandiga kanaler, som dämpar signalen jämnt. En teknisk term för denna typen av system är OFDM ("orthogonal frequency-division multiplexing"). Men med massiv MIMO leder ansamlingen av ett stort antal oberoende slumpmässiga källor ofta till en makroskopisk bild som verkar deterministisk. Detta kallas kanalhärdning. Vi undersöker huruvida denna härdande effekt förenklar den bearbetning som krävs utan att behöva dela upp den bredbandiga kanalen i flera icke-interfererande smalbandiga kanaler, det vill säga, utan att använda OFDM. Effekten av fasbrus i bredbandiga kanaler studeras även när en viss enkel linjär behandlingsteknik används. Eftersom det finns fler antennlement på basstationen undersöks det om det är mer fördelaktigt att använda en gemensam lokal oscillator eller flera oberoende lokala oscillatorer. Massiv MIMO kräver sidoinformation om utbredningskanalen för att på lämplig sätt kombinera de mottagna signalerna vid basstationen. I massiv MIMO förvärfas denna sidoinformation med hjälp av förutbestämda signaler som kallas piloter. Fasbruset roterar dock den mottagna signalen vilket leder till en snedvridning av de mottagna piloterna. De grundläggande gränserna för kvaliteten på sidoinformationen förvärfvad via piloter i Massive MIMO, när fasbrus är närvarande, undersöks också. Slutligen studeras effekten av fasbruset på massiv MIMO med OFDM.

Acknowledgments

During the period of my doctoral studies there have been many people that played key role in my development. First and foremost, I would like to express my deepest gratitude to my main supervisor, Prof. Erik G. Larsson, for his support and supervision during all this period. His persistence to work on well-defined and important problems was pivotal to the quality and impact of my research. His meticulous review of the solutions that I was proposing was constantly transforming and improving my output and was helping me identify new directions and potential contributions.

I am also heavily indebted to Dr. Saif Khan Mohammed, now Assistant Professor in IIT Delhi, who served as co-advisor for the first years of my doctoral studies. He has been tireless in explaining various problems of my proposed solutions and in clarifying fundamental, but difficult to grasp, notions from Information Theory. He has also been very persistent to teach me how I should express my ideas in my articles clearly and effectively.

During the last years of my doctoral studies Associate Professor Dr. Emil Björnson served as my co-advisor, to whom I have to extend my genuine appreciation for his help. His feedback has been immediate, concise and accurate in every issue that was coming up. His remarkable effectiveness helped me proceed forward and acquire a broader perspective of my research area. He has also assisted me to further improve my writing style.

At the same time, I had the opportunity to work in very competitive environment with many other knowledgeable and helpful senior researchers and professors, such as Mikael Olofsson and Danyo Danev. A special thanks should be extended to Dr. Eleftherios Karipidis, now with Ericsson Research, who has helped me both with his technical knowledge and expertise but also with his personal advice. I am also grateful to Prof. Michail Matthaiou, Queen's University, Belfast, UK, for his continuing support and useful advice since my master's studies in TU Munich. Among my doctoral colleagues I would like to express my special gratitude to Dr. Hien

Quoc Ngo, with whom I have spent many hours discussing solutions on problems and clarifying the details of his work, which helped my research efforts. Finally, many thanks to all the other colleagues that have passed from the corridor of Communication Systems and have enlightened me with their presentations and during personal discussions.

Last but not least, I would like to thank my parents and my sister as silent and indirect “co-authors” of this dissertation. Their contribution was their support and care for all my life and the opportunity they gave me to work in a peaceful and comfortable environment free of distractions and worries.

Linköping, April 2016
Antonios Pitarokoilis

Abbreviations

AGC	Automatic Gain Control
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BS	Base Station
CC	Constant Channel
CIR	Channel Impulse Response
CSI	Channel State Information
dB	decibel
DFT	Discrete Fourier Transform
DL	Downlink
DMC	Discrete Memoryless Channel
EVM	Error Vector Magnitude
FC	Fading Channel
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
GBC	Gaussian Broadcast Channel
IBI	Inter-Block Interference
IEEE	Institute of Electrical and Electronics Engineers
i.i.d.	Independent and Identically Distributed
ISF	Impulse Sensitivity Function
ISI	Inter-Symbol Interference
LLR	Log-Likelihood Ratio
LMMSE	Linear Minimum Mean Square Error
LO	Local Oscillator
LoS	Line-of-Sight
LPF	Low-Pass Filter
LTI	Linear Time Invariant
MAP	Maximum A Posteriori
MIMO	Multiple-Input Multiple-Output
MISO	Multiple-Input Single-Output

ML	Maximum Likelihood
MMSE	Minimum Mean Square Error
MRC	Maximum Ratio Combining
MRT	Maximum Ratio Transmission
MU	Multiuser
MUI	Multiuser Interference
NS	Non-Synchronous
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
PAM	Pulse Amplitude Modulation
PAPR	Peak-to-Average-Power-Ratio
PDP	Power Delay Profile
PLL	Phase-Lock Loop
PN	Phase Noise
PSK	Phase Shift Keying
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase Shift Keying
RF	Radio Frequency
RZF	Regularized Zero-Forcing
S	Synchronous
SBS	Symbol-by-Symbol
SER	Symbol Error Rate
SI	Side Information
SIMO	Single-Input Multiple-Output
SINR	Signal-to-Interference-plus-Noise Ratio
SISO	Single-Input Single-Output
SNR	Signal-to-Noise Ratio
SU	Single-User
TDD	Time Division Duplex
TR	Time-Reversal
UL	Uplink
UNF	Use-And-Forget
UT	User Terminal
WLAN	Wireless Local Area Network
ZF	Zero-Forcing

Contents

Abstract	v
Populärvetenskaplig Sammanfattning (in Swedish)	vii
Acknowledgments	ix
Abbreviations	xi
I Introduction	1
1 Motivation	3
2 Phase Noise in RF Oscillators	7
2.1 Macroscopic Manifestation of Phase Noise	7
2.2 Phase Noise Sources	9
2.3 Circuit-theoretic Modeling of Phase Noise	11
2.3.1 The Model by Leeson [25]	12
2.3.2 The Model by Hajimiri and Lee [13]	13
2.3.3 The Model by Demir, <i>et al.</i> [14]	14
3 Phase Noise in Communication Systems	17
3.1 The Complex Baseband Representation in the Presence of Phase Deviations	17
3.2 The AWGN Channel Impaired with Phase Noise	19
3.3 The Band-Limited Phase Noise Channel	21
3.4 The Wiener Phase Noise Model in Communications	24
4 Massive MIMO with Phase Noise Impairments	27
4.1 The MU-MIMO Uplink Channel	27
4.2 Massive MIMO	29
4.3 Massive MIMO with Phase Noise	31
4.3.1 Uplink Training	32
4.3.2 Data Transmission	32
4.4 Alternative Capacity Bound for Massive MIMO with Phase Noise	35

5	Contributions of the Dissertation	41
5.1	Included Papers	41
5.2	Not Included Papers	45
5.3	Future Research Directions	45
II	Included Papers	59
A	On the Optimality of Single-Carrier Transmission in Large-Scale Antenna Systems	61
1	Introduction	64
2	System Model	65
3	Achievable Sum-Rate	66
4	Simulation Results	71
B	Uplink Performance of Time-Reversal MRC in Massive MIMO Systems Subject to Phase Noise	77
1	Introduction	80
2	System Model	82
	2.1 Phase Noise Model	83
	2.2 Received Signal	84
3	Transmission Scheme and Receive Processing	84
	3.1 Channel Estimation	85
	3.2 Time-Reversal Maximum Ratio Combining (TR-MRC)	86
4	Achievable Sum-Rate	87
	4.1 Exact Analysis of Synchronous versus Non-Synchronous Operation for a Toy Channel Model	94
5	Asymptotic Results	95
6	Impact of Phase Noise Separately at the BS and at the User Terminals	97
	6.1 Special Case 1: Phase Noise Only at the UTs, $\sigma_\phi^2 = 0$	97
	6.2 Special Case 2: Phase Noise Only at the BS, ($\sigma_\phi^2 \neq 0$ and $\sigma_\theta^2 = 0$)	98
7	Numerical Examples	99
8	Conclusions	105
C	Achievable Rates of ZF Receivers in Massive MIMO with Phase Noise Impairments	117
1	Introduction	120
2	System Model	121
3	Transmission Scheme and ZF Receiver	122

3.1	LMMSE Channel Estimation	123
3.2	Zero-Forcing (ZF) Equalization	124
4	Achievable Rates	125
5	Results - Discussion	128
D ML Detection in Phase Noise Impaired SIMO Channels with		
Uplink Training		135
1	Introduction	138
2	System Model	141
2.1	Constant Channel (CC)	142
2.2	Fading Channel (FC)	143
3	Optimal Detectors	144
3.1	Implementation of (15) and (25)	148
3.2	The ML Detector for von Mises Phase Noise Increments	148
4	High SNR Analysis	150
4.1	High SNR Analysis for the Synchronous Operation .	150
4.2	High SNR Analysis for the CC-NS Case	152
4.3	High SNR Analysis for the FC-NS Case	154
5	Numerical Examples	156
5.1	Extension to Longer Data Intervals	161
6	Conclusions	163
7	Two-Slot Proofs	165
7.1	Proof of Proposition 1 for the CC-NS Case	165
7.2	Proof of Proposition 1 for the FC-NS Case	167
7.3	Proof of Proposition 2 for the CC-S Case	169
7.4	Proof of Proposition 2 for the FC-S Case	169
7.5	Proof of Proposition 3	169
8	T-slot Detectors	170
8.1	Decision Feedback Detector: CC-NS	170
8.2	"Better-than-optimal" Detector: CC-S	172
8.3	Decision Feedback Detector: FC-NS	173
8.4	"Better-than-optimal" Detector: FC-S	175
8.5	Numerical Examples	176
E Performance of the Massive MIMO Uplink with OFDM and		
Phase Noise		185
1	Introduction	188
2	System Model	189
3	Achievable Rates	191
4	Numerical Examples	194
5	Conclusion	196

Part I

Introduction

Chapter 1

Motivation

Arguably, the world has experienced a rapid growth of data traffic over wireless networks in the last decades. It is reported that the global mobile data traffic has increased by more than 400 million times over the past 15 years, from less than 10 GB per month in 2000 to 3.7 EB¹ per month at the end of 2015 [1]. This trend is not expected to cease any time soon. In fact it is predicted that it will accelerate in the years to come, as the traffic is expected to reach 30 EB per month by 2020 [1]. This demand is expected to originate not only from wireless data exchange by tablets and smartphones but also from the proliferation of new types of communication, such as machine-to-machine communication [2]. In short, any device that could collect and disseminate information via the wireless medium and could benefit from this data exchange, is expected to be equipped with wireless communication capabilities.

The straightforward way to satisfy the increase in traffic demand is to increase the frequency spectrum for the communication. However, there are various reasons that make this approach not attractive. First, the spectrum is a constrained natural resource. Within a fixed spectrum portion multiple communication services, military and civilian, must be accommodated. Already, most part of the available spectrum is allocated to various services and operators, so if we are to allocate more spectrum to mobile communications, spectrum portions from other services must be made available. This might be possible, but only to a very restricted extent, which is definitely not enough to fulfill the demand for the future mobile data traffic.

¹1 EB (ExaByte) is 1 billion (10^9) GB (GigaByte).

Also, certain frequency spectrum bands are not appropriate for wireless networks, due to unfavorable propagation conditions or high atmospheric attenuation. Finally, spectrum is a very expensive resource for mobile operators. It is clear that a more aggressive and spectrally efficient approach must be applied for the evolution of wireless networks. A more aggressive spectrum reuse can increase data rates within certain geographical areas of interest. This can be partly implemented with the deployment of small cells (including WiFi access points). However, small cells are less effective when dealing with high-mobility users and providing wide-area coverage.

The deployment of multiple antennas at the transmitter and receiver (multiple-input multiple-output (MIMO)) was shown to provide significant gains in the amount of information that can be communicated within a fixed frequency band. The gains scale linearly with the minimum of the transmit and receive antennas in the presence of rich scattering environment and when the channel is known at the receiver [3,4]. Simultaneously with the emergence of point-to-point MIMO, the concept of multi-user MIMO (MU-MIMO) was investigated, where a base station (BS) equipped with a handful of antennas communicates with a few non-cooperative users over the same time and frequency resource [5]. Early work on the topic includes [6–8]. Point-to-point MIMO provide the promised gains in rich scattering scenarios, but require large antenna separation in the presence of a strong Line-of-Sight (LoS) component to achieve the same performance. In contrast, MU-MIMO are more robust in LoS conditions. However, the adoption of MU-MIMO techniques in modern wireless communication standards has not been proportional to the available research literature due to the fact that the optimal transmission strategies are complex and require accurate channel state information (CSI) at the BS.

Massive MIMO [9], also known as Large Scale Antenna Systems, Very Large MIMO, Full-Dimension MIMO, proposes a new approach towards the practical implementation of ideas from MU-MIMO, where a BS with an unprecedentedly large number of BS antenna elements, M , serves simultaneously a few tens of single-antenna, non-cooperative users, K . When $M \gg K$ linear, low-complexity processing techniques are close to optimal, while at the same time, channel state acquisition can be made available at the BS via uplink training. This way the resources spent for channel training are proportional to the number of users and, thus, the design is scalable with M . Massive MIMO is shown to achieve substantial gains in spectral and radiated energy efficiency of future wireless networks [10,11].

Massive MIMO is not just an increase of the number of antenna elements at the BS and a corresponding adoption of linear processing strategies. This newly proposed physical layer can also have further implications on the design of wireless networks. Currently, most wireless standards use multicarrier techniques to transmit data over wireless channels. Even though multicarrier transmission is very likely to be part of next generation wireless networks, it would be useful to investigate whether with Massive MIMO there exist regimes or deployment scenarios where low complexity single-carrier transmission can provide comparable or even superior performance than currently used techniques. In particular, with single-carrier transmission low peak-to-average-ratio (PAPR) waveforms can be designed. With such waveforms, low quality, inexpensive, power efficient, non-linear power amplifiers can be used at each BS antenna element, which is important in Massive MIMO. In current systems expensive, power inefficient, highly linear amplifiers with a large power back-off are required to transmit signals with high PAPR. Multi-carrier signals are known to have a large PAPR. With single-carrier transmission, all the available bandwidth is allocated to a user, which simplifies the user scheduling. Further, the transmission when $M \gg K$ will be robust against intersymbol interference. Finally, single-carrier transmission is less sensitive to impairments, such as carrier phase instabilities, in comparison to multi-carrier techniques, such as orthogonal frequency division multiplexing (OFDM).

The deployment of economically viable Massive MIMO relies on the use of multiple inexpensive components that are likely to introduce distortions to the information signal. Hence, affordable Massive MIMO must be robust against hardware impairments that arise due to imperfections in the transmitter and the receiver circuits. Therefore, the study of the effect of hardware impairments, which is frequently treated as secondary in the study of conventional systems, is very important in Massive MIMO. One of the most important hardware impairments is phase noise, which is introduced in communication systems due to imperfections in the circuits of the local oscillators (LOs) at the transmitter and receiver chains. Imperfect LOs introduce random and time-varying rotations at the information signals. These rotations hinder the receiver to make the correct decisions about the transmitted signal based on the available noisy observations. Phase noise affects not only the data detection, but also the channel estimation. As Massive MIMO relies on the coherent combining of the received signals with estimated channel impulse responses in order to provide the promised gains, phase noise destroys the coherency between the channel estimates and the

effective channels at the time of decoding.

In this dissertation we investigate the performance of Massive MIMO with wideband transmission –single carrier and multi-carrier– and in the presence of phase noise impairments. Capacity lower bounds are derived for various scenarios of practical interest and scaling laws with respect to radiated energy efficiency and other performance measures of interest are stated. The effect of phase noise on detection with training based CSI is also studied rigorously by deriving the maximum likelihood (ML) detector. The introduction of the dissertation is structured as follows. In Chapter 2 the fundamental phenomena that give rise to phase noise in radio frequency oscillators are reviewed and the most important models are stated. In Chapter 3 the way that phase noise enters in communication systems is analyzed. In Chapter 4 the concept of Massive MIMO is briefly introduced and the basic steps for the analysis of phase noise impaired Massive MIMO systems are explained. In Chapter 5 the basic contributions of the dissertation are described and the included papers are listed.

Chapter 2

Phase Noise in RF Oscillators

In this chapter the issue of phase noise in radio frequency (RF) LOs is discussed. The purpose is to relate the models used in this dissertation and in communications literature, in general, to well-known, established models that correspond to the physical reality of contemporary LOs. The discussion will be concentrated on free-running LOs, which is the main assumption on the oscillator operation throughout the dissertation. Initially, the macroscopic effects of phase noise on periodic waveforms are described. The main noise sources in electronic circuits that eventually perturb the waveform at the output of an LO are briefly reviewed. Finally, three main models that have been widely used to explain the behavior of phase noise are described.

2.1 Macroscopic Manifestation of Phase Noise

The LOs are electronic circuits that are designed to produce an oscillating waveform with a specified angular frequency, ω_c . They are essential in wireless communications, since they are used to modulate the baseband signal to passband at the transmitter and to demodulate the received passband signal to baseband at the receiver. Ideally, the output voltage, $V_{\text{out}}(t)$, of a noiseless LO is a sinusoidal waveform that is perfectly stable in amplitude, frequency and phase, given by

$$V_{\text{out}}(t) = A \cos(\omega_c t + \phi). \quad (2.1)$$

However, the electronic components (active and passive) that are used to manufacture the LO circuits are impaired by noise. Consequently, the output of a real LO has amplitude and phase that are time-varying and the output voltage is given by

$$V_{\text{out}}(t) = (1 + a(t)) \cos(\omega_c(t + \alpha(t))), \quad (2.2)$$

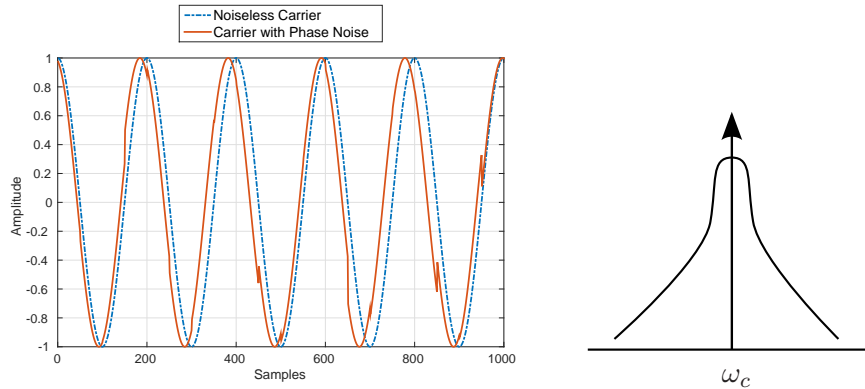
where $a(t)$ is the amplitude perturbation and $\alpha(t)$ is the timing noise. Due to the amplitude limiting mechanism that is present in practical LOs, e.g., as an automatic gain control (AGC) component [12], the amplitude perturbation, $a(t)$, gradually decays. Hence, the amplitude can be assumed to be constant. On the other hand, it has been shown that the noise $\alpha(t)$ persists and gives rise to a sequence of non-vanishing, random and varying timing offsets [13, 14].

In the time domain, phase noise becomes apparent by observing the zero-crossings of the oscillating waveform. The time interval between two consecutive zero-crossings from positive to negative values in this oscillating waveform, which is usually the voltage or the current at the output of the LO, should be exactly equal to the period of the LO waveform. However, in the presence of phase noise, these zero-crossings are slightly shifted, as shown in Fig. 2.1a. In the frequency domain, the power spectral density of the LO output is a Dirac impulse located at ω_c . In the presence of phase noise, however, the power spectral density of the LO output, assuming that it is wide-sense stationary, widens around the oscillation angular frequency, ω_c (Fig. 2.1b).

The effect of phase noise in digital modulation is a random phase rotation of the constellation points. In Fig. 2.2 the received symbols from a 16-QAM constellation [15] are plotted in the presence of phase noise when there is no additive thermal noise. The phase deviations in this figure are assumed to be zero-mean Gaussian¹ random variables with variance $10^{-2} \text{ rad}^2/\text{sample}$. It can be seen that the received symbols can be substantially rotated by phase noise. Consequently, the probability that a received signal is detected in error is high even when the desired signal power is significantly larger than the thermal noise power [16, 17]. The degradation due to phase noise is more detrimental for dense constellations, that convey information not only on the amplitude but also on the phase of the transmitted symbol.

Phase noise can introduce distortion not only to the transmitted/received signal but also to signals in adjacent frequency bands. At the receiver

¹The justification of this choice is deferred to Section 2.3.3 and Section 3.4.



(a) Phase noise in the time domain. The zero-crossings are shifted randomly from the nominal position in the presence of phase noise.

(b) Phase noise in the frequency domain. The ideal frequency spectrum widens due to phase noise.

Figure 2.1: Time-domain and frequency-domain manifestation of the output voltage from an LO with phase noise.

side, consider the demodulation of a weak passband signal with center frequency ω_1 , in the presence of a strong interferer in an adjacent frequency band centered at ω_2 , as shown in Fig. 2.3a. In the case of a noisy LO, significant interference can be introduced to the desired signal at ω_1 , from the strong interferer at ω_2 , shown as the shaded spectrum in Fig. 2.3b. This phenomenon is called *reciprocal mixing* [18]. At the transmitter side, a strong signal with a noisy LO at ω_2 , can introduce interference leakage to a weak neighboring signal at ω_1 (Fig. 2.4). Since the frequency bands around ω_1 and ω_2 are adjacent, the difference between the two carrier frequencies can be of the order of a few kHz. This imposes strict requirements so that the power spectral density of the noisy LOs should decay very sharply [18].

2.2 Phase Noise Sources

Noise is unavoidable in electronic circuits. Various random microscopic events that occur within a circuit and its components give rise to fluctuations and disturbances of the current and the voltage at the output of the circuit. The random motion of electrons within a resistor result in a voltage,

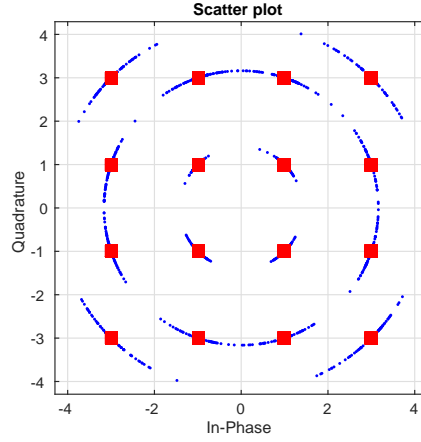


Figure 2.2: 16-QAM constellation rotation in the presence of phase noise and the absence of additive thermal noise.

$e_n(t)$, along the resistor that has a non-zero mean-square value even in the absence of current flow through the resistor [19]. A resistor of R Ohm has a mean square noise density of

$$\overline{e_n^2} = 4kTR\Delta f, \quad (2.3)$$

over a bandwidth of Δf Hz, where $k \approx 1.38 \times 10^{-23}$ Joule/Kelvin is the Boltzmann constant and T is the temperature measured in Kelvin. This type of noise is referred to as *thermal noise* due to its dependence on the absolute temperature, T . The mean square noise density is independent of the frequency, f , for a very wide spectrum of frequencies, which implies that thermal noise is white for most of the frequencies of interest. Thermal noise is also known as Johnson-Nyquist noise, due to the scientists that first measured and explained the phenomenon [20,21].

Shot noise appears in devices such as diodes or bipolar transistors in the presence of a constant average current flow of I_{DC} Ampère. These devices have a potential barrier that the charge carriers have to cross [19]. Due to the discrete type of the microscopic electron motions and the randomness in the time instants when these motions occur, fluctuations around I_{DC} appear. A shot noise source has a mean square current noise at a bandwidth of Δf Hz given by

$$\overline{i_n^2} = 2qI_{DC}\Delta f, \quad (2.4)$$

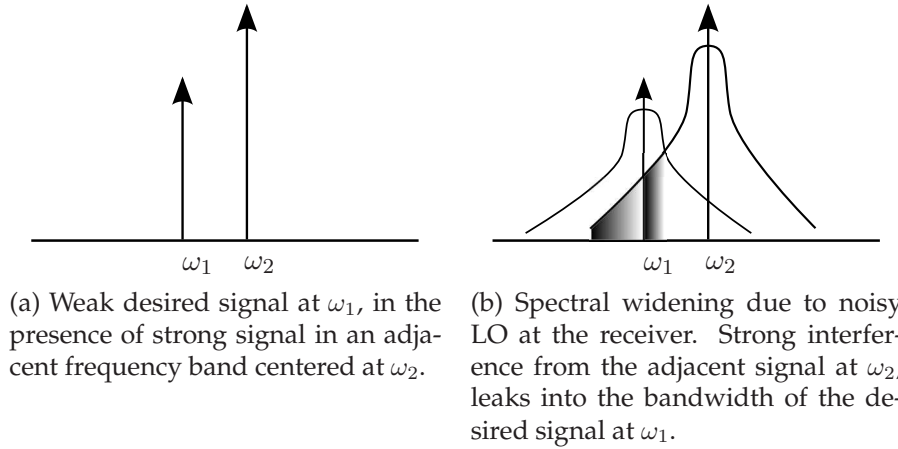


Figure 2.3: Reciprocal mixing due to noisy LO at the receiver [18].

where $q \approx 1.6 \times 10^{-19}$ Coulomb is the electron charge. The mean square current noise i_n^2 is independent of the frequency, f , which implies that shot noise is also white. The description and explanation of shot noise is attributed to W. Schottky [22].

Another type of noise that is present in physical systems is *flicker noise*. It has been observed in the fluctuations of various physical phenomena, spanning from biology, to electronics and astrophysics. This type of noise is colored and has power spectral density that scales inverse-proportionally with the frequency. For this reason, flicker noise is also called $1/f$ -noise [23, 24]. The physical mechanism that generates flicker noise is still debatable in the physics community and its description involves empirical parameters [19].

2.3 Circuit-theoretic Modeling of Phase Noise

In this section we briefly review three of the most important models that have been used to explain the behavior of the output of noisy LOs. The models are presented in the chronological order that they appeared and the main contribution on the understanding and modeling of phase noise is concisely explained.

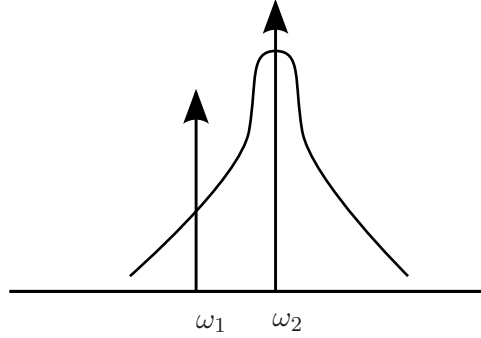


Figure 2.4: Interference leakage at the transmitter due to strong signal with high phase noise spectrum at ω_2 in the bandwidth of the desired signal at ω_1 .

2.3.1 The Model by Leeson [25]

The most common measure of the output of a noisy LO, usually a voltage or a current, is the single-sided power spectral density defined by [13]

$$\mathcal{L}(\Delta\omega) \triangleq 10 \log_{10} \left[\frac{\mathcal{P}(\omega_c + \Delta\omega, 1 \text{ Hz})}{\mathcal{P}_c} \right], \quad (2.5)$$

where $\mathcal{P}(\omega_c + \Delta\omega, 1 \text{ Hz})$ is the single-sided power measured at frequency band of width 1 Hz located at an offset of $\Delta\omega$ from the carrier, ω_c , and \mathcal{P}_c is the power of the carrier. The unit of $\mathcal{L}(\Delta\omega)$ is decibels below carrier per Hertz, $(\text{dBc}/\text{Hz})^2$ [26]. The single-sided power spectral density at the output of a noisy LO as a function of the offset, $\Delta\omega$, from the nominal oscillation frequency, ω_c , has, in general, a shape similar to the one shown in Fig. 2.5. Three different regions are apparent, namely, i) the small offset region, where $\mathcal{L}(\Delta\omega)$ drops as $1/f^3$, ii) the medium offset region, where $\mathcal{L}(\Delta\omega)$ drops as $1/f^2$ and iii) the large offset region, where $\mathcal{L}(\Delta\omega)$ reaches a constant level. Leeson in 1966 [25] attempted to explain this spectrum using linear time-invariant system theory. The derived spectrum is given by

$$\mathcal{L}(\Delta\omega) = 10 \log_{10} \left[\frac{2FkT}{P_{\text{sig}}} \left\{ 1 + \left(\frac{\omega_c}{2Q\Delta\omega} \right)^2 \right\} \left(1 + \frac{\Delta\omega_1/f^3}{|\Delta\omega|} \right) \right], \quad (2.6)$$

²The careful reader will observe that the “per Hertz” should apply to the quantities inside the logarithm. However, in practice the unit is used as here, with the implicit understanding that Hz should refer to the argument of the logarithm rather than the logarithm itself.

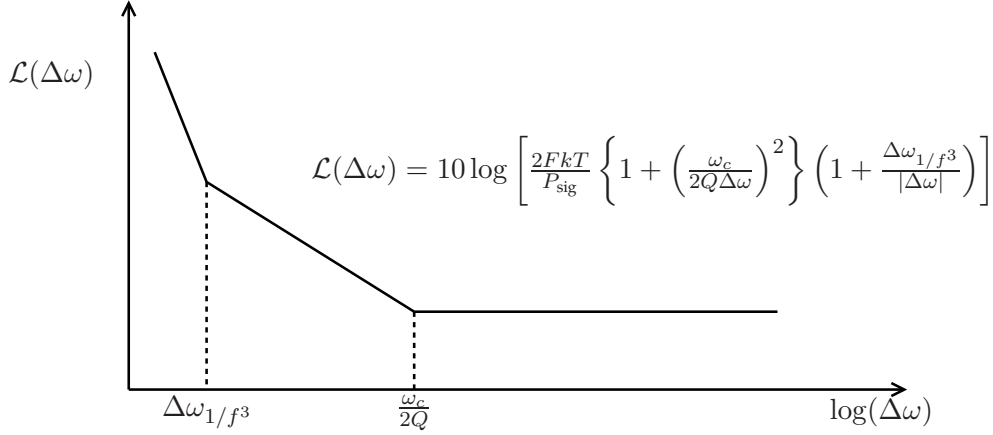


Figure 2.5: Single-sided power spectral density of a noisy LO output as predicted by Leeson [25].

where $\Delta\omega$ is the frequency offset from the carrier, k is the Boltzmann constant, T is the temperature in Kelvin, P_{sig} is the signal power, Q is the resonator³, F is the device excess noise number [13] and $\Delta\omega_{1/f^3}$ is the corner frequency offset between the $1/f^3$ region and the $1/f^2$ region. Apart from matching the predicted spectrum at the output of a noisy LO, Leeson's model also reveals some useful rules of thumb. For instance, an LO with high Q , i.e., good oscillator, will have a better phase noise performance. Leeson's model, despite its heuristic nature, has been a major reference for many years.

2.3.2 The Model by Hajimiri and Lee [13]

Leeson's model has certain drawbacks. For instance, F is an empirical parameter that cannot be predicted before the design of the LO. Further, the value of $\Delta\omega_{1/f^3}$ in (2.6) must also be determined empirically in practice. The reason for these deviations from reality is the fact that LOs are neither linear nor time-invariant systems. Hence, the Leeson approach, which relied on these assumptions, did not properly reflect the physical reality.

Hajimiri and Lee [13, 26] showed that LOs are time-varying systems and dropped the time-invariance assumption. They represented the noise

³This is a parameter that shows the quality of the oscillator and is defined as the ratio of the energy stored in the oscillator's reactive elements over the energy dissipated [19].

sources in a general LO as current impulses and they argued that the resulting perturbations on the LO output can be decomposed to pure amplitude and pure phase perturbations. While the amplitude perturbations go to zero due to the amplitude limiting mechanisms in the circuit, the phase perturbations do not. Further, they argued that the impulse response of the varying phase, $\phi(t)$, at the LO output to a current noise impulse (current-to-phase transfer function) is still linear but time-varying. They introduced special functions that they called *impulse sensitivity functions* (ISFs), which contain information on the sensitivity of the output LO waveform to circuit noise impulses. In general, ISFs depend on the oscillating waveform and are usually determined numerically. The authors used the theory of linear time-varying systems to derive the spectrum at the output of a noisy LO. The derived model gives a more accurate and less heuristic explanation of the power spectral density of a noisy LO output. Therefore, it helps designers identify the dominant sources of phase noise and take them into consideration in their design.

2.3.3 The Model by Demir, *et al.* [14]

The models in Sections 2.3.1 and 2.3.2 explain the phenomenon of phase noise in terms of the output power spectral density rather than the time-domain statistical characterization of the phase noise process. Such a description is more useful to the designers of LOs and less to communications engineers. An asymptotic statistical characterization of the phase noise process in the time domain is given by Demir, *et al.* [14]. They revisited the conjecture of Hajimiri and Lee that the perturbations can be decomposed to pure amplitude and pure phase perturbations and studied LOs as dynamical systems, i.e., systems that can be described by a differential equation in the presence of some perturbation

$$\frac{d}{dt}x(t) = f(x(t)) + B(x(t))b(t), \quad (2.7)$$

where $x(t)$ is the state of the system—typically a vector of voltages and/or currents along capacitors and inductors of the LO—at time t , $f(\cdot)$ is some—not necessarily linear—function and $B(x(t))b(t)$ is a small state-dependent perturbation. They assumed that there is a stable non-trivial periodic solution, $x_s(t)$, (“orbit” or “limit cycle” [27]) to the unperturbed system, i.e., (2.7) without the term $B(x(t))b(t)$, and they showed that with a small perturbation the orbit changes to $x_s(t + \alpha(t)) + y(t)$, where $y(t)$ is an amplitude perturbation that eventually disappears, whereas the time shift, $\alpha(t)$,

persists. Hence, they showed that the behavior of $\alpha(t)$ demonstrates the accumulative nature of phase noise. They also proceeded to show that, asymptotically in t and when the circuit noise sources are white (thermal and shot noise), the characteristic function of $\alpha(t)$ is that of a Gaussian random variable and that the autocorrelation function of the time shift $\alpha(t)$ is given by

$$\mathbb{E}[\alpha(t)\alpha(t + \tau)] = m^2 + c \min(t, t + \tau), \quad (2.8)$$

where m and c are constants. The asymptotic statistical characterization of $\alpha(t)$ in (2.8) implies that the increments $\alpha(t) - \alpha(t + \tau)$ are Gaussian and have variance that is proportional to $|\tau|$ [28]. These properties correspond to a continuous-time *Brownian motion* or *Wiener process* [29, Section 37] and justify a model that is widely used in communications, the *Wiener phase noise model*. The fact that phase noise can be described in the time domain with a very compact characterization that depends on a single scalar, c , makes the Wiener model very attractive for study of phase noise in the fields of Information Theory and Detection & Estimation Theory. A more detailed analysis of this model in a communications setup is deferred to Section 3.4. Demir extended the analysis to include white and colored noise sources (i.e., flicker noise), however, the expressions attained are much more complicated and involve the integration over the spectra of the colored noise sources [30, Lemma 7.1].

Chapter 3

Phase Noise in Communication Systems

In the following the impairment of phase noise is introduced in suitable communication system models. The discussion starts with the complex baseband representation of a real passband signal as it passes through a communication system with phase deviations at the transmitter and the receiver. Subsequently, the single-input single-output (SISO) phase noise impaired additive white Gaussian noise (AWGN) channel is discussed and based on the continuous-time system model, a discrete-time approximation is derived. The band-limited SISO channel with phase noise impairments at the transmitter and the receiver is also introduced and discrete-time approximations are derived. The chapter concludes with the description of the Wiener phase noise model, which has been motivated by Section 2.3.3, and its connection to the communication-theoretic system models is described.

3.1 The Complex Baseband Representation in the Presence of Phase Deviations

Let two sequences $a(t)$ and $b(t)$ of pulse amplitude modulation (PAM) symbols to be transmitted over a channel located at a frequency, f_c . The channel is assumed to be noiseless and does not introduce any other distortion to

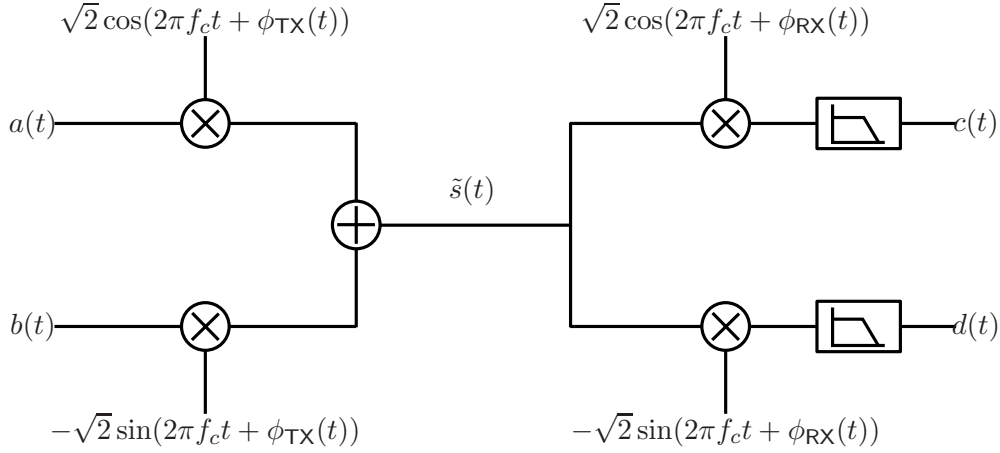


Figure 3.1: Passband representation of a system with phase deviations at the transmitter and the receiver.

the transmitted signal. However, we assume that phase noise distortions, $\phi_{\text{TX}}(t)$ and $\phi_{\text{RX}}(t)$, are introduced due to noisy LOs at the transmitter and the receiver, respectively. The real passband signal after the modulation is given by

$$\begin{aligned}\tilde{s}(t) &= a(t)\sqrt{2}\cos(2\pi f_c t + \phi_{\text{TX}}(t)) - b(t)\sqrt{2}\sin(2\pi f_c t + \phi_{\text{TX}}(t)) \\ &= \Re \left\{ (a(t) + jb(t))e^{j\phi_{\text{TX}}(t)}\sqrt{2}e^{j2\pi f_c t} \right\}.\end{aligned}\quad (3.1)$$

Hence, the complex baseband equivalent representation of the transmitted signal in the presence of phase deviations at the transmitter is given by $s_b(t) = (a(t) + jb(t))e^{j\phi_{\text{TX}}(t)}$.

At the receiver, the in-phase component, $c(t)$, is given by

$$\begin{aligned}c(t) &= \text{LPF} \left\{ \tilde{s}(t)\sqrt{2}\cos(2\pi f_c t + \phi_{\text{RX}}(t)) \right\} \\ &= a(t)\cos(\phi_{\text{TX}}(t) - \phi_{\text{RX}}(t)) - b(t)\sin(\phi_{\text{TX}}(t) - \phi_{\text{RX}}(t)),\end{aligned}\quad (3.2)$$

where $\text{LPF} \{ \cdot \}$ is the operation of a low-pass filter, which filters out the frequency component at $2f_c$. Similarly, the quadrature component is given by

$$\begin{aligned}d(t) &= \text{LPF} \left\{ \tilde{s}(t)\sqrt{2}\sin(2\pi f_c t + \phi_{\text{RX}}(t)) \right\} \\ &= a(t)\sin(\phi_{\text{TX}}(t) - \phi_{\text{RX}}(t)) + b(t)\cos(\phi_{\text{TX}}(t) - \phi_{\text{RX}}(t)).\end{aligned}\quad (3.3)$$

The received complex baseband signal is

$$y(t) = c(t) + jd(t) = (a(t) + jb(t))e^{j\phi_{\text{TX}}(t)}e^{-j\phi_{\text{RX}}(t)}. \quad (3.4)$$

In the case of narrowband SISO systems, the baseband equivalent representation in (3.4) reduces to $y(t) = (a(t) + jb(t))e^{j\theta(t)}$, where only the difference $\theta(t) = \phi_{\text{TX}}(t) - \phi_{\text{RX}}(t)$ is important. However, in MU-MIMO systems the distinction between the transmit and receive phase deviations is essential. Consider for example a system with two non-cooperative single-antenna transmitters and a single-antenna receiver. The two effective channels from the transmitters to the receiver will be both affected by the same receive phase noise process but from independent transmit phase noise processes. This must be reflected by the selected channel representation.

3.2 The AWGN Channel Impaired with Phase Noise

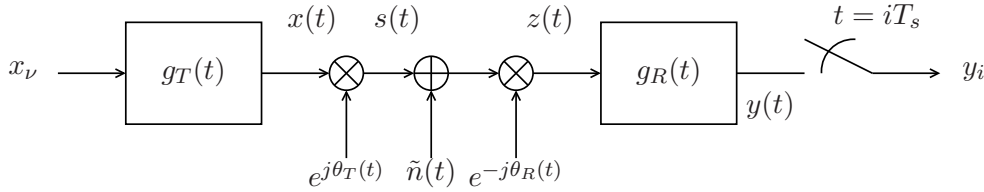


Figure 3.2: The continuous-time additive white Gaussian (AWGN) channel with phase noise.

Based on the complex baseband equivalent representation of phase noise impaired SISO systems from Section 3.1, the phase noise impaired AWGN channel is studied in continuous time and a discrete time approximation is derived. Let the waveform to be transmitted, $x(t)$, be

$$x(t) = \sum_{\nu=1}^N x_{\nu} g_T(t - \nu T_s), \quad (3.5)$$

where T_s is the symbol interval, $\{x_1, \dots, x_N\}$ is the sequence of transmitted symbols selected from a fixed constellation [15] and $g_T(t)$ is the impulse response of the transmit pulse shaping filter. The actually transmitted waveform is disturbed by the transmit phase noise process, $\theta_T(t)$:

$$s(t) = e^{j\theta_T(t)} x(t) = e^{j\theta_T(t)} \sum_{\nu=1}^N x_{\nu} g_T(t - \nu T_s). \quad (3.6)$$

The channel introduces to the transmitted signal, $s(t)$, additive noise $\tilde{n}(t)$, which is a white Gaussian random process. At the receiver the signal after the downconversion by the LO at the receiver, $z(t)$, is given by

$$z(t) = \sum_{\nu=1}^N x_{\nu} e^{-j(\theta_R(t) - \theta_T(t))} g_T(t - \nu T_s) + n(t). \quad (3.7)$$

Due to the circular symmetry of the $\tilde{n}(t)$ noise process, $n(t) = e^{-j\theta_R(t)} \tilde{n}(t)$ and $\tilde{n}(t)$ have the same second order statistical characterization. Further, it is clear that $\theta_T(t)$ and $\theta_R(t)$ are observed only through their difference. Hence, in what follows we note $\theta(t) \triangleq \theta_T(t) - \theta_R(t)$. At the receiver, the signal is filtered by a linear time-invariant (LTI) filter with impulse response, $g_R(t)$, which is matched to $g_T(t)$. The output of the filter $y(t)$ is given by

$$y(t) = \int_{-\infty}^{+\infty} g_R(\tau) z(t - \tau) d\tau = \sum_{\nu=1}^N x_{\nu} \int_{-\infty}^{+\infty} e^{j\theta(t-\tau)} g_T^*(-\tau) g_T(t - \tau - \nu T_s) d\tau + \int_{-\infty}^{+\infty} g_R(\tau) n(t - \tau) d\tau. \quad (3.8)$$

The mathematical manipulation of the derived continuous-time model appears to be formidable. In addition, the progress of digital electronics motivates the discretization of the continuous-time model in (3.8). For this purpose, the received signal, $y(t)$, is sampled at regular intervals equal to the symbol interval, T_s . The sample y_i at time $t = iT_s$ is given by

$$y_i = y(t = iT_s) = \sum_{\nu=1}^N x_{\nu} \int_{-\infty}^{+\infty} e^{j\theta(iT_s - \tau)} g_T^*(-\tau) g_T(iT_s - \tau - \nu T_s) d\tau + n_i \quad (3.9)$$

A common assumption is that phase noise is constant within a symbol interval, but it does change between two consecutive symbol intervals. Under this assumption and if the cascade of the transmit and receive filters $g_T(t)$ and $g_R(t)$ satisfies the Nyquist criterion for intersymbol interference [15] the following simple symbol-sampled phase noise channel model follows immediately

$$y_i = x_i e^{j\theta_i} + n_i, \quad (3.10)$$

where the noise samples n_i are jointly Gaussian and uncorrelated, hence statistically independent [31, Section 3.3.3]. Under these conditions, the sequence y_1, \dots, y_N is sufficient statistics for the detection of x_1, \dots, x_N . The

model (3.10) has been extensively studied under various assumptions on the statistics of the phase sequence, $\{\theta_i\}$. The arguments used above for the derivation of the discrete-time symbol-sampled phase noise model are in general well-known and can be found in literature either implicitly [32] or more explicitly [33,34].

3.3 The Band-Limited Phase Noise Channel

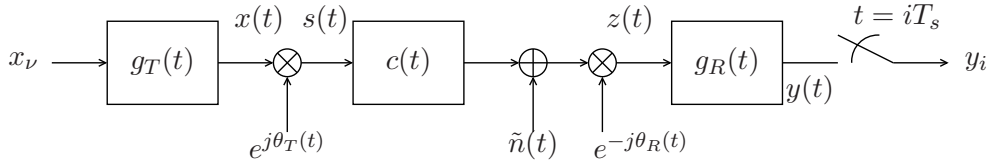


Figure 3.3: The continuous-time band-limited channel with phase noise.

Here we extend the discussion to the case of a band-limited channel with complex baseband equivalent impulse response $c(t)$, when phase noise is introduced at the transmitter and the receiver, as shown in Fig. 3.3. We assume that the signal to be transmitted is given by

$$x(t) = \sum_{\nu=1}^N x_{\nu} g_T(t - \nu T_s).$$

The transmitter signal, which is distorted by transmit phase noise, is given by

$$s(t) = e^{j\theta_T(t)} x(t) = e^{j\theta_T(t)} \sum_{\nu=1}^N x_{\nu} g_T(t - \nu T_s). \quad (3.11)$$

The received signal after the filtering from the channel with impulse response $c(t)$, the distortion by additive Gaussian noise, $\tilde{n}(t)$, and the demodulation by the imperfect LO at the receiver is given by

$$\begin{aligned} z(t) &= e^{-j\theta_R(t)} \int_{-\infty}^{+\infty} c(\tau_1) s(t - \tau_1) d\tau_1 + e^{-j\theta_R(t)} \tilde{n}(t) \\ &= e^{-j\theta_R(t)} \int_{-\infty}^{+\infty} c(\tau_1) \left(e^{j\theta_T(t-\tau_1)} \sum_{\nu=1}^N x_{\nu} g_T(t - \tau_1 - \nu T_s) \right) d\tau_1 + e^{-j\theta_R(t)} \tilde{n}(t) \\ &= \sum_{\nu=1}^N x_{\nu} \int_{-\infty}^{+\infty} e^{-j(\theta_R(t) - \theta_T(t-\tau_1))} c(\tau_1) g_T(t - \tau_1 - \nu T_s) d\tau_1 + n(t) \end{aligned}$$

At the output of the receive filter, $g_R(t)$, the signal is given by

$$\begin{aligned}
y(t) &= \int_{-\infty}^{+\infty} g_R(\tau_2) z(t - \tau_2) d\tau_2 \\
&= \sum_{\nu=1}^N x_\nu \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-j(\theta_R(t-\tau_2) - \theta_T(t-\tau_2-\tau_1))} \\
&\quad \times g_R(\tau_2) c(\tau_1) g_T(t - \tau_2 - \tau_1 - \nu T_s) d\tau_1 d\tau_2 + \int_{-\infty}^{+\infty} g_R(\tau_2) n(t - \tau_2) d\tau_2
\end{aligned} \tag{3.12}$$

It is clear that a discrete-time model derived from the continuous-time representation is necessary. A rigorous argumentation that can lead to a tractable discrete-time description appears to be difficult. Often an approximation similar to the one applied to discretize (3.9) is invoked. That is, the phase noise processes are assumed to be significantly more narrowband than the phase-noise-free overall pulse shape $h(t) = (g_R * c * g_T)(t)$. Under this assumption, the transmit and receive phase noise processes are treated as one and the discrete-time model is given by

$$y_i = e^{-j\theta[i]} \sum_{l=0}^{L-1} h_l x_{i-l} + n_i, \tag{3.13}$$

where $h_l = (g_R * c * g_T)(lT_s)^1$ and $\theta[i]$ is the sampled equivalent phase noise process. Such an approach has been adopted by many authors, such as [28, 35–37], which are studies of single-input single-output (SISO) orthogonal frequency division multiplexing (OFDM) systems.

There are two outstanding issues with the discrete-time model in (3.13). First, the phase noise processes at the transmitter and the receiver are treated as an equivalent phase noise process. This is a problem in MIMO and multi-user systems, where independent phase noise processes at different antenna elements or at different users must be appear separately. Second, the phase noise processes at the transmitter and the receiver, $\theta_T(t)$ and $\theta_R(t)$, in (3.12) are distinguishable as $e^{j\theta_T(t)}$ is convolved with the propagation channel, $c(t)$, as well as with the receive filter, $g_R(t)$. In contrast, $e^{-j\theta_R(t)}$ is only convolved with $g_R(t)$. We seek for a discrete-time approximation that will reflect this property of the continuous-time representation.

¹Here, it has also been assumed that the overall pulse shape, $h(t)$, gradually decays to negligible values, so we can truncate the sampled impulse response to L symbol-spaced channel taps.

We assume that the transmit and receive filters, $g_T(\cdot)$ and $g_R(\cdot)$, are ideal *integrate-and-dump* filters, i.e., their impulse responses are given by

$$g_T(t) = g_R(t) = \begin{cases} \frac{1}{\sqrt{T_s}}, & 0 \leq t \leq T_s \\ 0, & \text{elsewhere.} \end{cases} \quad (3.14)$$

Then the continuous-time model in (3.12) is expressed as

$$\begin{aligned} y(t) &= \sum_{\nu=1}^N x_\nu \int_0^{T_s} \int_{t-\tau_2-\nu T_s-T_s}^{t-\tau_2-\nu T_s} e^{-j(\theta_R(t-\tau_2)-\theta_T(t-\tau_2-\tau_1))} \frac{1}{T_s} c(\tau_1) d\tau_1 d\tau_2 \\ &\quad + \int_0^{T_s} \frac{1}{\sqrt{T_s}} n(t-\tau_2) d\tau_2 \end{aligned} \quad (3.15)$$

With equidistant symbol-spaced sampling at $t = iT_s$, the discrete-time model derived from the continuous-time model in (3.15) is given by

$$\begin{aligned} y_i &= y(t = iT_s) \\ &= \sum_{\nu=1}^N x_\nu \int_0^{T_s} \int_{(i-\nu-1)T_s-\tau_2}^{(i-\nu)T_s-\tau_2} e^{-j(\theta_R(iT_s-\tau_2)-\theta_T(iT_s-\tau_2-\tau_1))} \frac{1}{T_s} c(\tau_1) d\tau_1 d\tau_2 \\ &\quad + \int_0^{T_s} \frac{1}{\sqrt{T_s}} n(iT_s - \tau_2) d\tau_2. \end{aligned} \quad (3.16)$$

We invoke the piecewise-continuous assumption on the phase noise processes successively on $\theta_T(\cdot)$ and $\theta_R(\cdot)$, i.e., we assume that the phase noise process has not changed substantially within an interval of duration T_s , and we also define

$$\tilde{c}(t) \triangleq \frac{1}{T_s} \int_{t-T_s}^t c(\tau) d\tau. \quad (3.17)$$

and

$$n_i \triangleq \int_0^{T_s} \frac{1}{\sqrt{T_s}} n(iT_s - \tau) d\tau. \quad (3.18)$$

Then, the discrete-time approximation is given by

$$\begin{aligned} y_i &\approx \sum_{\nu=1}^N x_\nu e^{j\theta_T(\nu T_s)} \int_0^{T_s} e^{-j\theta_R(iT_s-\tau_2)} \tilde{c}((i-\nu)T_s - \tau_2) d\tau_2 + n_i \\ &\approx \sum_{\nu=1}^N x_\nu e^{-j\theta_R((i-1)T_s)} e^{j\theta_T(\nu T_s)} \int_0^{T_s} \tilde{c}((i-\nu)T_s - \tau_2) d\tau_2 + n_i. \end{aligned}$$

With the additional definition and truncation of the discrete-time impulse response

$$h_l \triangleq \begin{cases} \int_0^{T_s} \tilde{c}(lT_s - \tau) d\tau, & \text{for } 0 \leq l \leq L - 1 \\ 0, & \text{elsewhere} \end{cases} \quad (3.19)$$

a discrete-time approximation of the continuous-time model in (3.12) is given by

$$y_i = e^{-j\theta_R[i]} \sum_{l=0}^{L-1} h_l e^{j\theta_T[i-l]} x_{i-l} + n_i. \quad (3.20)$$

The system model in (3.20) is appealing since the phase noise processes at the transmitter and the receiver appear explicitly and the convolution effect of the propagation channel on $\theta_T(t)$ is also given as a discrete-time convolutional sum. Consequently, it has been extensively used in the literature to model phase noise in frequency selective channels [38–40]. However, in contrast to the discrete-time model in (3.10), a detailed derivation of the model (3.20), as the one given here, appears to be absent in the relevant literature.

3.4 The Wiener Phase Noise Model in Communications

In the Sections 3.1, 3.2 and 3.3 we described the way that phase noise appears in standard communication system models [38,41–44]. In this section we present a widely studied model on phase noise in the fields of Information Theory and Detection & Estimation Theory. Based on the discussion of Section 2.3.3, the phase noise process, $\theta(t) = \omega_c \alpha(t)$, in (3.8) is modeled as a continuous-time Wiener process, i.e.,

$$\theta(t) - \theta(t_0) = \int_{t_0}^t w(\tau) d\tau, \quad (3.21)$$

where t_0 is the start of the observation of the process and $w(t)$ is a real white Gaussian process with $\mathbb{E}[w(t)] = 0$ and $\mathbb{E}[w(t_1)w(t_2)] = \omega_c^2 c \delta(t_1 - t_2)$. Of particular interest is the autocorrelation and power spectral density of the

process $\zeta(t) = e^{j(\theta(t)-\theta(t_0))}$. It is straightforward from the definitions of $\zeta(t)$ and $\theta(t)$ to show that

$$\begin{aligned} R_\zeta(\tau) &= \mathbb{E}[\zeta(t)\zeta^*(t+\tau)] = \mathbb{E}\left[e^{j(\theta(t)-\theta(t_0))}e^{-j(\theta(t+\tau)-\theta(t_0))}\right] \\ &= \mathbb{E}\left[e^{j\omega_c(\alpha(t)-\alpha(t+\tau))}\right] = e^{-\frac{\omega_c^2}{2}|\tau|}. \end{aligned} \quad (3.22)$$

The fact that $R_\zeta(\tau)$ is only function of the relative time lag τ implies the stationarity of the process $\zeta(t)$. The corresponding power spectral density is

$$\begin{aligned} S_\zeta(f) &= \int_{-\infty}^{+\infty} R_\zeta(\tau)e^{-j2\pi f\tau} d\tau = \int_{-\infty}^{+\infty} e^{-\frac{\omega_c^2}{2}|\tau|}e^{-j2\pi f\tau} d\tau \\ &= \frac{1}{\pi} \frac{\pi f_c^2 c}{\pi^2 f_c^4 c^2 + f^2} = \frac{1}{\pi} \frac{\beta/2}{(\beta/2)^2 + f^2}. \end{aligned} \quad (3.23)$$

The shape of the spectrum in (3.23) is Lorentzian and $\beta \triangleq 2\pi f_c^2 c$ is the double-sided 3 dB bandwidth of the process. Observe that this shape is an accurate description of the medium offset region of $\mathcal{L}(\Delta\omega)$ in Fig. 2.5. However, it does not exhibit the $1/f^3$ behavior for small offsets from carrier. This is due to the fact that the Wiener phase noise characterization is valid for LOs with only white circuit noise sources (thermal and shot noise) and the $1/f^3$ behavior is attributed to colored circuit noise sources, e.g., flicker noise.

In the discrete time, the continuous-time Wiener process is sampled at a frequency $1/T_s$ to yield the discrete-time Wiener process, where

$$\theta_{i+1} = \theta_i + w_i, \quad (3.24)$$

where $\theta_i = \theta(iT_s)$ and

$$w_i = \int_{iT_s}^{(i+1)T_s} w(\tau) d\tau.$$

From the properties of the continuous-time Wiener process, the increments w_i are independent identically distributed (i.i.d.) zero mean real Gaussian random variables with variance $\sigma_w^2 = 2\pi\beta T_s$, i.e., $w_i \sim \mathcal{N}_{\mathbb{R}}(0, \sigma_w^2)$. In information-theoretic studies it is often assumed that the initial phase θ_0 is uniform in $[0, 2\pi)$, i.e., $\theta_0 \sim \mathcal{U}[0, 2\pi)$, and then the process $\{\theta_i\}$ is stationary [45].

Chapter 4

Massive MIMO with Phase Noise Impairments

In this chapter, the concept of MU-MIMO systems is introduced and the basic system model is described for the single-cell case. The introduction of Massive MIMO follows, where the main differences with conventional MU-MIMO systems are explained. The outline of the achievable sum-rate analysis in Massive MIMO is briefly presented. Subsequently, phase noise impairments are introduced into the basic Massive MIMO model. Through simple yet illustrative examples, the fundamental differences in the achievable sum-rate analysis between the phase noise case and phase-noise-free case are reviewed. The capacity lower bounds used in this dissertation are presented in detail, their desirable characteristics are mentioned and the potential use of alternative capacity lower bounds is discussed.

4.1 The MU-MIMO Uplink Channel

Consider a BS with M antenna elements serving simultaneously K single-antenna, non-cooperative users¹ within a geographical area, called *cell*. We study the direction of communication where the users send data to the BS; this is called *uplink*. Narrowband transmission is assumed, i.e., the effect

¹The single-antenna per user assumption is not essential and is done here for simplicity. However, the non-cooperation between users is fundamental.

of the propagation through the channel between a user and a BS antenna is represented by a single complex scalar coefficient. We denote the channel gain from the k -th user to the m -th BS antenna by $h_{m,k}$ and the channel vector from the k -th user to the BS by $\mathbf{h}_k \triangleq [h_{1,k}, \dots, h_{M,k}]^T$ and assume that it is a circularly symmetric complex Gaussian random vector, i.e., $\mathbf{h}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{I}_M)$. If the k -th user transmits the symbol x_k , which is subject to an average power constraint, $\mathbb{E}[|x_k|^2] \leq 1$, then the received vector at the BS is given by

$$\mathbf{y} = \sqrt{\rho} \sum_{k=1}^K \mathbf{h}_k x_k + \mathbf{w}_k, \quad (4.1)$$

where $\mathbf{y} \in \mathbb{C}^M$ is an M -dimensional complex vector and \mathbf{w}_k is additive white Gaussian noise, distributed as $\mathbf{w}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{I}_M)$. If $\mathbb{E}[|h_{m,k}|^2] = 1$, $\forall m, k$ and only one user transmits, then ρ is the expected signal-to-noise-ratio (SNR) at each BS antenna. Important research results on the limits of communication on the uplink channel are already available in the literature [46–49]. The downlink, which is the communication direction from the BS to the users, is also important and has been extensively studied [50–57].

The concept of MU-MIMO systems was in itself a paradigm shift with respect to point-to-point single-user (SU) MIMO. In point-to-point SU-MIMO both the BS and the user terminal are equipped with multiple antenna elements and different users are scheduled in orthogonal channels. In the presence of a strong LoS path between the user and the BS or of strong spatial correlation, the number of streams that can be multiplexed in the SU-MIMO case can drop to one, offering only a small performance gain in comparison to the case where the BS and the user have a single antenna. However, MU-MIMO is more robust under LoS conditions, since two users are likely to be spatially separated and cases of strong correlation among users can be resolved by appropriate multi-user scheduling. Also, due to the single-antenna assumption for the user terminals, the complexity is moved to the BS and the user terminals can be small and energy efficient. However, MU-MIMO has also drawbacks. Accurate knowledge of the channels, \mathbf{h}_k , is required to reap the gains of MU-MIMO. This knowledge acquisition is costly, particularly in the downlink direction [5]. In addition, the optimal transmit and receive strategies in MU-MIMO, where the number of BS antennas, M , is approximately equal to the number of single-antenna users, K , are non-linear with very high complexity and simpler linear schemes often perform poorly.

4.2 Massive MIMO

Massive MIMO proposes a further shift in the paradigm of cellular systems design [58,59]. A Massive MIMO BS is equipped with an unprecedentedly large number of antenna elements, M , a few hundreds or even thousands, and serves simultaneously a few tens of non-cooperative, single-antenna users, K . In this operational regime, where $M \gg K$, the normalized propagation vector channels from different users tend to become asymptotically orthogonal. This is called *favorable propagation*. Further, the normalized channel norms tend to be very close to their statistical mean. This phenomenon is called *channel hardening*. With favorable propagation and channel hardening, linear transmit and receive strategies are close to optimal. Similar to conventional MU-MIMO, accurate knowledge of the channels at the BS side is required in Massive MIMO in order to reap the promised gains. This knowledge is typically acquired via uplink training. With this choice, the length of the required training interval increases with the number of terminals and not with the number of BS antennas. Finally, the BS can use the channel knowledge acquired via uplink training to detect the signals received in the uplink.

The concept of *coherence interval* is key in the study of wireless channels, and of Massive MIMO in particular. A coherence interval is the block, of say τ_c channel uses, during which the channel remains approximately constant. In the present exposition of Massive MIMO the coherence interval is split into an uplink training interval of $\tau_p \geq K$ channel uses and an uplink data interval of $\tau_d = \tau_c - \tau_p$ channel uses². During training, each user transmits a deterministic training sequence $\sqrt{\tau_p}\psi_k$ that is orthogonal to the training sequences of the other users, i.e., $\psi_k^H \psi_{k'} = \delta(k - k')$, where $\delta(\cdot)$ is the Kronecker delta. The received vectors during training are given by

$$\mathbf{Y}_p = \sqrt{\rho\tau_p} \sum_{k=1}^K \mathbf{h}_k \psi_k^T + \mathbf{W}_p \quad (4.2)$$

$$\mathbf{Y}_p \psi_k^* = \sqrt{\rho\tau_p} \mathbf{h}_k + \mathbf{W}_p \psi_k^*, \quad (4.3)$$

where $\mathbf{Y}_p = [\mathbf{y}[1], \dots, \mathbf{y}[\tau_p]]$ and $\mathbf{W}_p = [\mathbf{w}[1], \dots, \mathbf{w}[\tau_p]]$. The BS forms an estimate, $\hat{\mathbf{h}}_k$, for the channel vector, \mathbf{h}_k , based on $\mathbf{Y}_p \psi_k^*$ in order to detect the

²In Massive MIMO part of the coherence interval is allocated also for downlink data transmission and, possibly, for downlink pilots [60], but in this introductory exposition the focus is on the uplink.

information symbols transmitted from the k -th user, $x_k[\tau_p + 1], \dots, x_k[\tau_c]$. Information symbols from different users are assumed to be independent. During the data transmission interval, the BS uses the estimated channels³ with maximum ratio combining (MRC)⁴ to detect the transmitted data symbol, $x_k[i]$, at the i -th channel use of the data interval, i.e.,

$$\hat{x}_k[i] = \hat{\mathbf{h}}_k^H \mathbf{y}[i] = \sqrt{\rho} \sum_{k'=1}^K \hat{\mathbf{h}}_k^H \mathbf{h}_{k'} x_{k'}[i] + \hat{\mathbf{h}}_k^H \mathbf{w}[i]. \quad (4.4)$$

If $\mathbb{E} \left[x_k^* \hat{\mathbf{h}}_k^H \mathbf{w}[i] \middle| \hat{\mathbf{h}}_k \right] = 0$ and $\mathbb{E} \left[\left(\hat{\mathbf{h}}_k^H \mathbf{h}_k x_k \right)^* \hat{\mathbf{h}}_k^H \mathbf{w}[i] \middle| \hat{\mathbf{h}}_k \right] = 0$, a lower bound (achievable rate), R_k^{SI} , on the maximum mutual information between $x_k[i]$ and $\hat{x}_k[i]$ conditioned on $\hat{\mathbf{h}}_k$, $I(x_k[i]; \hat{x}_k[i] | \hat{\mathbf{h}}_k)$ is obtained by

$$\begin{aligned} \max_{p_{X_k[i]}(x_k[i])} I(x_k[i]; \hat{x}_k[i] | \hat{\mathbf{h}}_k) &\geq R_k^{\text{SI}} \\ &= \mathbb{E} \left[\log_2 \left(1 + \frac{\rho \left| \mathbb{E} \left[\hat{\mathbf{h}}_k^H \mathbf{h}_k \middle| \hat{\mathbf{h}}_k \right] \right|^2}{\rho \sum_{k'=1}^K \mathbb{E} \left[\left| \hat{\mathbf{h}}_k^H \mathbf{h}_{k'} \right|^2 \middle| \hat{\mathbf{h}}_k \right] - \rho \left| \mathbb{E} \left[\hat{\mathbf{h}}_k^H \mathbf{h}_k \middle| \hat{\mathbf{h}}_k \right] \right|^2 + \left\| \hat{\mathbf{h}}_k \right\|^2} \right) \right]. \end{aligned} \quad (4.5)$$

The maximization is over all the input densities $p_{X_k[i]}(x_k[i])$ that satisfy the power constraint $\mathbb{E} \left[|x_k[i]|^2 \right] \leq 1$. We will use the term *side-information bound* for R_k^{SI} since it explicitly uses the side information acquired via uplink training both to process the received signal vector $\mathbf{y}[i]$ and to decode the information symbol $x_k[i]$ from the output, $\hat{x}_k[i]$, of the receive processing filter. A more tractable but looser bound, R_k^{UNF} , on the maximum mutual information between $x_k[i]$ and $\hat{x}_k[i]$, $I(x_k[i]; \hat{x}_k[i])$, is given by

$$\begin{aligned} \max_{p_{X_k[i]}(x_k[i])} I(x_k[i]; \hat{x}_k[i]) &\geq R_k^{\text{UNF}} \\ &= \log_2 \left(1 + \frac{\rho \left| \mathbb{E} \left[\hat{\mathbf{h}}_k^H \mathbf{h}_k \right] \right|^2}{\rho \sum_{k'=1}^K \mathbb{E} \left[\left| \hat{\mathbf{h}}_k^H \mathbf{h}_{k'} \right|^2 \right] - \rho \left| \mathbb{E} \left[\hat{\mathbf{h}}_k^H \mathbf{h}_k \right] \right|^2 + \mathbb{E} \left[\left\| \hat{\mathbf{h}}_k \right\|^2 \right]} \right). \end{aligned} \quad (4.6)$$

We will use the term *use-and-forget bound* for R_k^{UNF} since $\hat{\mathbf{h}}_k$ is only used to process the received vector $\mathbf{y}[i]$ but this information is not explicitly used to

³The users have no knowledge of the estimates $\hat{\mathbf{h}}_k$, hence the data symbols $x_k[i]$ are independent of the estimates $\hat{\mathbf{h}}_k$.

⁴Some other linear or with low complexity processing scheme, such as linear minimum mean square error (LMMSE), is also possible.

decode the information symbol $x_k[i]$ from $\hat{x}_k[i]$. A detailed derivation and information-theoretic justification of these bounds can be found in [61]. We note that these bounds first appeared in [62–64] and since then they have been extensively used in the study of Massive MIMO systems [10,11,65–67].

4.3 Massive MIMO with Phase Noise

In this section the basic single-cell system model in (4.1) is augmented with the effect of transmit and receive phase noise and the methodology for the derivation of achievable rates is outlined and explained. For simplicity, the exposition here is restricted to a single user, however, all the fundamental phenomena that appear due to phase noise are readily revealed. The treatment of more complicated models can be found in the papers that follow. At the i -th channel use of the coherence interval, the received signal at the m -th BS antenna is given by

$$y_m[i] = \sqrt{\rho} e^{-j\theta_m[i]} h_m e^{j\phi[i]} x[i] + w_m[i], \quad (4.7)$$

based on the analysis leading to (3.20) (with $L = 1$). The processes $\theta_m[\cdot]$, $\phi[\cdot]$, h_m , $x[\cdot]$ and $w_m[\cdot]$ are mutually independent. However, BS phase noise processes from the m_1 -th and m_2 -th BS antennas, $\theta_{m_1}[\cdot]$ and $\theta_{m_2}[\cdot]$, respectively, can be arbitrarily dependent. We consider two particular operations; the *synchronous* operation where $\theta_1[\cdot] \equiv \dots \equiv \theta_M[\cdot]$ and the *non-synchronous* operation where the $\theta_m[\cdot]$ are mutually independent along the BS antennas. The synchronous operation models a centralized deployment where one LO provides the carrier waveform to all the BS antennas. In contrast, the non-synchronous operation models a distributed deployment where separate LOs are used for every BS antenna. All the phase noise processes are assumed to be discrete-time Wiener processes as defined in Section 3.4. The variances of the phase noise increments for $\theta_m[\cdot]$ and $\phi[\cdot]$ are σ_θ^2 and σ_ϕ^2 , respectively. The matrix-vector formulation is given by

$$\mathbf{y}[i] = \sqrt{\rho} \mathbf{\Theta}[i] \mathbf{h} e^{j\phi[i]} x[i] + \mathbf{w}[i], \quad (4.8)$$

where $\mathbf{y}[i] \in \mathbb{C}^M$, $\mathbf{\Theta}[i] \triangleq \text{diag}\{e^{-j\theta_1[i]}, \dots, e^{-j\theta_M[i]}\}$ for the non-synchronous operation and $\mathbf{\Theta}[i] \triangleq e^{-j\theta[i]} \mathbf{I}_M$ for the synchronous operation.

As it has been outlined in Section 4.2, a coherence interval of τ_c channel uses is assumed with τ_p channel uses for uplink training and τ_d channel

uses for uplink data transmission. The propagation channel, \mathbf{h} , remains constant for the coherence interval, changes to an independent realization between different coherence intervals and the fading process is assumed to be stationary and ergodic.

4.3.1 Uplink Training

For simplicity, the training sequence consists of τ_p ones. At the l -th channel use of the training interval the received signal, $\mathbf{y}[l]$, is given by

$$\mathbf{y}[l] = \sqrt{\rho}\Theta[l]\mathbf{h}e^{j\phi[l]} + \mathbf{w}[l]. \quad (4.9)$$

It is clear that, even though \mathbf{h} remains constant during the coherence interval, the phase noise processes evolve at symbol rate. Hence, \mathbf{h} has to be estimated in the presence of the time-varying disturbance due to phase noise. We select to use the following estimate

$$\hat{\mathbf{h}} = \frac{1}{\tau_p\sqrt{\rho}} \sum_{l=1}^{\tau_p} \mathbf{y}[l] = \frac{1}{\tau_p} \sum_{l=1}^{\tau_p} \Theta[l]\mathbf{h}e^{j\phi[l]} + \frac{1}{\tau_p\sqrt{\rho}} \sum_{l=1}^{\tau_p} \mathbf{w}[l]. \quad (4.10)$$

This particular estimate is the ML estimate of \mathbf{h} given the received vectors $\mathbf{y}[1], \dots, \mathbf{y}[\tau_p]$ in the absence of phase noise and when $\mathbf{w}[l]$ is Gaussian. It is also linear and with low complexity, which are important requirements for channel estimation in Massive MIMO. The use of the minimum mean square error (MMSE) estimate is another choice. Various fundamental papers on Massive MIMO [9–11] have used this estimate. In (4.9) the coefficients of the effective vector channels are rotations of the propagation channel, \mathbf{h} , due to the presence of phase noise. Hence, the calculation of the MMSE estimate is formidable. The calculation of the linear MMSE estimate is straightforward, however, its use does not make the derivation of the achievable rates substantially easier with respect to the “quasi”-ML estimate of (4.10) nor does it change the fundamental insights drawn from the final result.

4.3.2 Data Transmission

The BS uses the channel knowledge acquired via training in (4.10) to perform MRC on the received vectors, $\mathbf{y}[i]$, during data transmission. The detected symbol, $\hat{x}[i]$, based on processing of the received vector $\mathbf{y}[i]$ is given

by

$$\hat{x}[i] = \hat{\mathbf{h}}^H \mathbf{y}[i] = \sqrt{\rho} \hat{\mathbf{h}}^H \Theta[i] \mathbf{h} e^{j\phi[i]} x[i] + \hat{\mathbf{h}}^H \mathbf{w}[i]. \quad (4.11)$$

Observe that even though $\hat{\mathbf{h}}$ provides important information about \mathbf{h} , this is in the form of a superposition of various versions of \mathbf{h} with elements that are randomly rotated by phase noise. Further, the effective channel gains at the i -th channel use, $\Theta[i] \mathbf{h} e^{j\phi[i]}$, are also rotated versions of the propagation channel, \mathbf{h} . Therefore, it appears formidable to exploit all the information available by the channel estimate, $\hat{\mathbf{h}}$, further than its use in (4.11), as it was done in (4.5). An extension of the approach leading to (4.6) is henceforth outlined. By expanding the estimated channel, $\hat{\mathbf{h}}$, the detected symbol, $\hat{x}[i]$, in (4.11) is written as

$$\hat{x}[i] = \sqrt{\rho} A[i] x[i] + \frac{1}{\tau_p} \sum_{l=1}^{\tau_p} \mathbf{w}^H[l] \Theta[l] \mathbf{h} e^{j\phi[l]} x[i] + \hat{\mathbf{h}}^H \mathbf{w}[i], \quad (4.12)$$

where

$$A[i] \triangleq \frac{1}{\tau_p} \sum_{l=1}^{\tau_p} \mathbf{h}^H \Theta^H[l] \Theta[l] \mathbf{h} e^{j(\phi[i] - \phi[l])}. \quad (4.13)$$

Observe that $A[i]$ is random due to its dependence on phase noise. However, its moments can be computed. Hence, we can express (4.12) as

$$\hat{x}[i] = \sqrt{\rho} \mathbb{E}[A[i]] x[i] + \text{EN}[i] \quad (4.14)$$

where the effective noise term $\text{EN}[i]$ is given by

$$\text{EN}[i] = \sqrt{\rho} (A[i] - \mathbb{E}[A[i]]) x[i] + \frac{1}{\tau_p} \sum_{l=1}^{\tau_p} \mathbf{w}^H[l] \Theta[l] \mathbf{h} e^{j\phi[l]} x[i] + \hat{\mathbf{h}}^H \mathbf{w}[i]. \quad (4.15)$$

Observe that $\mathbb{E}[A[i]] x[i]$ and $\text{EN}[i]$ are uncorrelated but they are not necessarily independent. With $x_k[i] \sim \mathcal{N}_{\mathbb{C}}(0, 1)$ the worst case additive uncorrelated noise, with respect to mutual information, is circularly symmetric complex Gaussian with 0 mean and variance equal to the variance of $\text{EN}[i]$. Hence, a lower bound on the mutual information $I(x[i]; \hat{x}[i])$ is given by

$$I(x[i]; \hat{x}[i]) \geq R^{\text{UNF}}[i] = \log_2 \left(1 + \frac{\rho |\mathbb{E}[A[i]]|^2}{\text{VAR}(\text{EN}[i])} \right), \quad (4.16)$$

where

$$\mathbb{E}[A[i]] = \frac{M}{\tau_p} \sum_{l=1}^{\tau_p} e^{-\frac{\sigma_\phi^2 + \sigma_\theta^2}{2}|i-l|}. \quad (4.17)$$

For the particular cases of synchronous and non-synchronous operation the expression $\text{VAR}(\text{EN}[i])$ is

$$\text{VAR}(\text{EN}[i]) = \rho \left(M(M+1)t_1 - |\mathbb{E}[A[i]]|^2 \right) + Mt_1 + \frac{M}{\tau_p \rho} + \frac{M}{\tau_p}, \quad (4.18)$$

for the synchronous operation and

$$\text{VAR}(\text{EN}[i]) = \rho \left(2Mt_1 + M(M-1)t_2 - |\mathbb{E}[A[i]]|^2 \right) + Mt_1 + \frac{M}{\tau_p \rho} + \frac{M}{\tau_p}, \quad (4.19)$$

for the non-synchronous operation, where the terms t_1 and t_2 are given by

$$t_1 = \frac{1}{\tau_p^2} \sum_{l_1=1}^{\tau_p} \sum_{l_2=1}^{\tau_p} e^{-\frac{\sigma_\phi^2 + \sigma_\theta^2}{2}|l_1-l_2|} \quad (4.20)$$

$$t_2 = \frac{1}{\tau_p^2} \sum_{l_1=1}^{\tau_p} \sum_{l_2=1}^{\tau_p} e^{-\frac{\sigma_\phi^2}{2}|l_1-l_2|} e^{-\frac{\sigma_\theta^2}{2}|i-l_1|} e^{-\frac{\sigma_\theta^2}{2}|i-l_2|}. \quad (4.21)$$

The overall effective achievable rate is given then by

$$R^{\text{UNF}} = \frac{1}{\tau_c} \sum_{i=\tau_p+1}^{\tau_c} R^{\text{UNF}}[i]. \quad (4.22)$$

The result in (4.22) has certain appealing properties. First, it is in closed-form. This implies that no time-consuming and complicated numerical studies are required to predict the performance of phase-noise-impaired Massive MIMO. The effects on the achievable rate with respect to parameters, such as the number of the BS antennas, M , the variance of the phase noise increments, σ_ϕ^2 and σ_θ^2 and the length of the training interval, τ_p , appear explicitly in the derived achievable rate. This facilitates the statement of various scaling laws with respect to the individual parameters, e.g., how does the achievable rate change as a function of M . Also, connections between different parameters can be revealed, e.g., how much can ρ be reduced with an increase of M , while a fixed desired achievable rate is guaranteed.

The achievable rate in (4.22) corresponds to a realistic scenario, i.e., an explicit and implementable channel estimation scheme is used to acquire CSI, the acquired CSI is used for data detection with low complexity, linear reception and the effect of the finite coherence interval is made explicit. The methodology provides insights for different deployment scenarios. Here, the synchronous and non-synchronous operations have been considered but similar analysis can be performed for arbitrary hybrid scenarios, where a single LO controls a group of antenna elements. Further, there is a plethora of results for phase-noise-free Massive MIMO, which are derived based on similar methodology [10,66,68], as outlined in Section 4.2. Quantifying the achievable rate loss due to phase noise in Massive MIMO is now straightforward by comparison to the existing results. Finally, the methodology applied here can be extended to more complex scenarios, e.g., multiple users, multiple cells, and also to more complex system models that include other hardware impairments [69,70].

4.4 Alternative Capacity Bound for Massive MIMO with Phase Noise

In the previous section the main capacity lower bound used in the dissertation was explained and justified. In this section we explore a special case where another achievable rate can be derived, based on the analysis leading to (4.5). That is, in this bound the acquired channel knowledge is exploited both for MRC detection and for symbol decoding. Even though, the approach seems difficult to be extended to realistic single-cell Massive MIMO systems with phase noise, its derivation and comparison with the used bound provides important insights on the gap in terms of achievable rate given by (4.16) and (4.22). We note that for the derivation of (4.16) and (4.22), part of the information provided by the estimated channel was not used in the decoding process.

Consider the same setup as in Section 4.3 with $\tau_p = 1$. In this case, the received signal during channel training is given by

$$\mathbf{y}[0] = \sqrt{\rho}\mathbf{h} + \mathbf{w}[0] \quad (4.23)$$

where the initial phase noise rotations have been absorbed into the propagation channel \mathbf{h} . This is possible without change in the statistics of \mathbf{h}

due to the circular symmetry of the distribution of \mathbf{h} . The BS calculates the MMSE estimate based on the observation $\mathbf{y}[0]$, i.e.,

$$\hat{\mathbf{h}} = \frac{\sqrt{\rho}}{1 + \rho} \mathbf{y}[0] \quad (4.24)$$

Due to the fact that the channel estimate, $\hat{\mathbf{h}}$, and the channel estimation error, $\tilde{\mathbf{h}}$, are jointly Gaussian and uncorrelated, they are also statistically independent. During data transmission, with MRC processing the detected symbol $\hat{x}[i]$ at the i -th channel use is given by

$$\hat{x}[i] = \hat{\mathbf{h}}^H \mathbf{y}[i] = \sqrt{\rho} \hat{\mathbf{h}}^H \Theta[i] \mathbf{h} e^{j\phi[i]} x[i] + \hat{\mathbf{h}}^H \mathbf{w}[i] \quad (4.25)$$

$$\begin{aligned} &= \sqrt{\rho} \mathbb{E} \left[\hat{\mathbf{h}}^H \Theta[i] \mathbf{h} e^{j\phi[i]} \middle| \hat{\mathbf{h}} \right] x[i] \\ &+ \underbrace{\sqrt{\rho} \left(\hat{\mathbf{h}}^H \Theta[i] \mathbf{h} e^{j\phi[i]} - \mathbb{E} \left[\hat{\mathbf{h}}^H \Theta[i] \mathbf{h} e^{j\phi[i]} \middle| \hat{\mathbf{h}} \right] \right)}_{\triangleq \text{EN}[i]} x[i] + \hat{\mathbf{h}}^H \mathbf{w}[i] \end{aligned} \quad (4.26)$$

An ergodic achievable rate is given by

$$R^{\text{SI}}[i] = \mathbb{E} \left[\log_2 \left(1 + \frac{\rho \left| \mathbb{E} \left[\hat{\mathbf{h}}^H \Theta[i] \mathbf{h} e^{j\phi[i]} \middle| \hat{\mathbf{h}} \right] \right|^2}{\rho \text{VAR} \left(\hat{\mathbf{h}}^H \Theta[i] \mathbf{h} e^{j\phi[i]} \middle| \hat{\mathbf{h}} \right) + \mathbb{E} \left[\left| \hat{\mathbf{h}}^H \mathbf{w}[i] \right|^2 \middle| \hat{\mathbf{h}} \right]} \right) \right]. \quad (4.27)$$

The above expression can be further simplified with the calculation of the conditional expectations $\mathbb{E} \left[\hat{\mathbf{h}}^H \Theta[i] \mathbf{h} e^{j\phi[i]} \middle| \hat{\mathbf{h}} \right] = e^{-\frac{\sigma_\phi^2 + \sigma_\theta^2}{2} i} \|\hat{\mathbf{h}}\|^2$ and $\mathbb{E} \left[\left| \hat{\mathbf{h}}^H \mathbf{w}[i] \right|^2 \middle| \hat{\mathbf{h}} \right] = \|\hat{\mathbf{h}}\|^2$. The inner expectations are conditioned on $\hat{\mathbf{h}}$ and with respect to all the remaining sources of randomness and the outer expectation is with respect to $\hat{\mathbf{h}}$. The expression $\text{VAR} \left(\hat{\mathbf{h}}^H \Theta[i] \mathbf{h} e^{j\phi[i]} \middle| \hat{\mathbf{h}} \right)$ evaluates to

$$\text{VAR} \left(\hat{\mathbf{h}}^H \Theta[i] \mathbf{h} e^{j\phi[i]} \middle| \hat{\mathbf{h}} \right) = \left(1 - e^{-(\sigma_\phi^2 + \sigma_\theta^2) i} \right) \|\hat{\mathbf{h}}\|^4 + \frac{1}{1 + \rho} \|\hat{\mathbf{h}}\|^2 \quad (4.28)$$

and

$$\begin{aligned} \text{VAR} \left(\hat{\mathbf{h}}^H \Theta[i] \mathbf{h} e^{j\phi[i]} \middle| \hat{\mathbf{h}} \right) &= \left(1 - e^{-\sigma_\theta^2 i} \right) \sum_{m=1}^M |\hat{h}_m|^4 + \frac{1}{1 + \rho} \|\hat{\mathbf{h}}\|^2 \\ &+ e^{-\sigma_\theta^2 i} \left(1 - e^{-\sigma_\phi^2 i} \right) \|\hat{\mathbf{h}}\|^4 \end{aligned} \quad (4.29)$$

for the synchronous and non-synchronous operation, respectively. This is a side-information bound that extends the result in (4.5) to Massive MIMO with phase noise.

The corresponding use-and-forget bound can also be derived as outlined in the Section 4.3 and in this case is given by

$$R^{\text{UNF}}[i] = \log_2 \left(1 + \frac{\rho \left| \mathbb{E} \left[\hat{\mathbf{h}}^H \Theta[i] \mathbf{h} e^{j\phi[i]} \right] \right|^2}{\rho \text{VAR} \left(\hat{\mathbf{h}}^H \Theta[i] \mathbf{h} e^{j\phi[i]} \right) + \mathbb{E} \left[\left| \hat{\mathbf{h}}^H \mathbf{w}[i] \right|^2 \right]} \right), \quad (4.30)$$

where the evaluation of the expectations yields $\mathbb{E} \left[\hat{\mathbf{h}}^H \Theta[i] \mathbf{h} e^{j\phi[i]} \right] = \frac{\rho}{\rho+1} M e^{-\frac{\sigma_\phi^2 + \sigma_\theta^2}{2} i}$ and $\mathbb{E} \left[\left| \hat{\mathbf{h}}^H \mathbf{w}[i] \right|^2 \right] = \frac{\rho}{\rho+1} M$. The variance term is evaluated for the synchronous operation to

$$\rho \text{VAR} \left(\hat{\mathbf{h}}^H \Theta[i] \mathbf{h} e^{j\phi[i]} \right) = \rho \left(\frac{\rho}{1+\rho} \right)^2 M^2 \left(1 - e^{-(\sigma_\phi^2 + \sigma_\theta^2) i} \right) + \frac{\rho^2}{\rho+1} M \quad (4.31)$$

and for the non-synchronous operation to

$$\begin{aligned} \rho \text{VAR} \left(\hat{\mathbf{h}}^H \Theta[i] \mathbf{h} e^{j\phi[i]} \right) &= \rho M \frac{\rho}{\rho+1} \frac{2\rho+1}{\rho+1} \\ &+ \rho M \left(\frac{\rho}{\rho+1} \right)^2 e^{-\sigma_\theta^2 i} \left((M-1) - M e^{-\sigma_\phi^2 i} \right). \end{aligned} \quad (4.32)$$

In Figs. 4.1, 4.2 and 4.3 the achievable rates based on (4.27) and (4.30) are plotted for the sake of comparison. In all cases, the SI bound is better than the UNF bound. This is expected since the SI bound uses explicitly the estimated channel both for the data detection and data decoding. In contrast, the UNF bound in (4.30) uses the estimated channel knowledge only during data detection and “forgets” it during data decoding. In Fig. 4.1 the achievable rates at the first channel use, i.e., $i = 1$, are plotted for $M = 100$, $\sigma_\phi^2 = \sigma_\theta^2 = 10^{-4}$ and $\tau_d = 100$ as a function of ρ , [dB]. The bounds are identical in the low ρ regime, but the performance difference is substantial at high ρ . However, both bounds, SI and UNF, saturate at high ρ . In Fig. 4.2 the achievable rates (4.27) and (4.30) are plotted for $M = 500$ and $i = 100$. It is observed that the gap between the SI and UNF has decreased significantly. This is partly due to the larger M and partly due to the partial loss

of the coherency between the estimated channel $\hat{\mathbf{h}}$ and the effective channel $\Theta[i]\mathbf{h}e^{j\phi[i]}$. At the 100-th data channel use the phase noise processes have evolved substantially so that omitting the information on $\hat{\mathbf{h}}$ during data decoding does not have a significant impact on the achievable rate performance. In Fig. 4.3 the average achievable rates over the coherence interval (calculated similarly to (4.22) using (4.27) and (4.30)) are plotted for $M = 100$, $\tau_d = 500$ and $\sigma_\phi^2 = \sigma_\theta^2 = 10^{-3}$ and the bounds appear to be very close to each other.

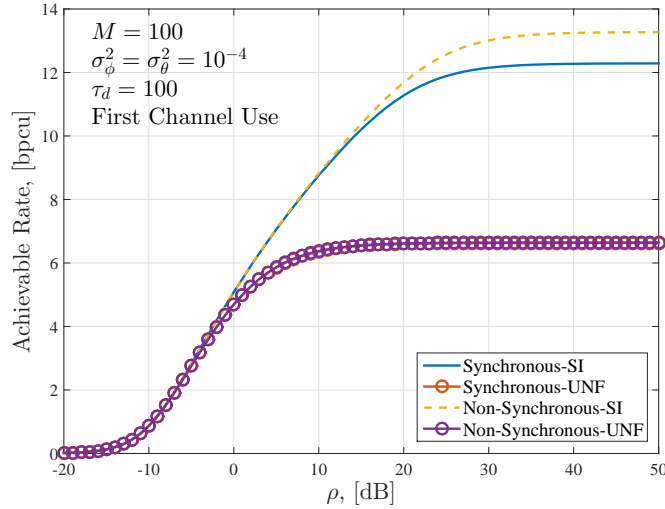


Figure 4.1: Comparison of the achievable rates in (4.27) and (4.30) for $i = 1$, $M = 100$, $\sigma_\phi^2 = \sigma_\theta^2 = 10^{-4}$ and $\tau_d = 100$ as a function of ρ , [dB].

In summary, in the examples it was observed that the SI bound yields better performance than the UNF bound. However, it also has several drawbacks. First, it is not in closed form and in particular for the non-synchronous operation the derivation of a closed-form expression appears to be formidable. Hence, we have to rely on Monte-Carlo simulations. The approach is difficult to be generalized for more practical scenarios, e.g., multi-user scenarios or longer training sequences. In particular, in the presence of multiple users the channel estimates are expected to be more noisy, hence the gain in achievable rate by exploiting the explicit form of the channel estimate, if this is possible, is likely to be very small. Also, the two bounds are far from each other only at high SNR, for small M , small phase noise variance, $\sigma_\phi^2, \sigma_\theta^2$, and for the initial channel uses of the coherence interval. For larger

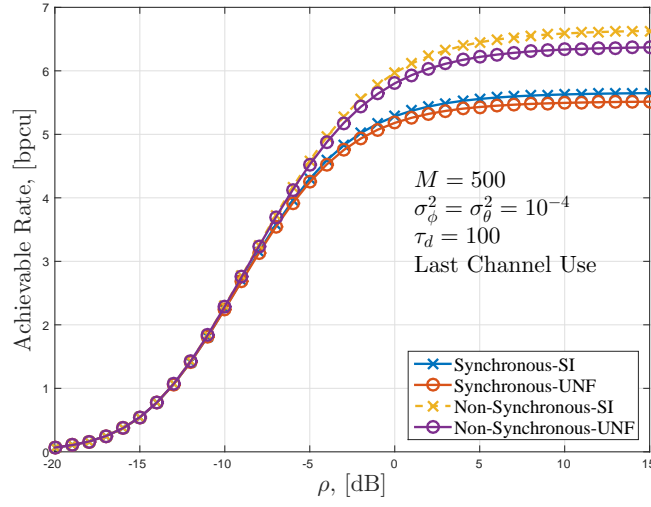


Figure 4.2: Comparison of the achievable rates in (4.27) and (4.30) for $i = 100$, $M = 500$, $\sigma_\phi^2 = \sigma_\theta^2 = 10^{-4}$ and $\tau_d = 100$ as a function of ρ , [dB].

M , larger phase noise variances and longer coherence intervals, the performance difference between the SI bound and UNF bound is very small even in the scenario examined here.

The sum-capacity lower bounds presented in this dissertation were derived similarly to the approach outlined in Section 4.3. Based on Section 4.4 there is evidence that they are very close to the sum-capacity lower bound obtained if the estimated CSI was used not only for processing of the received signal but also for decoding of the information symbol. At the same time the bounds summarize concisely the influence of multiple parameters on the achievable rate performance of Massive MIMO with phase noise, without the need to resort to time-consuming Monte-Carlo simulations.

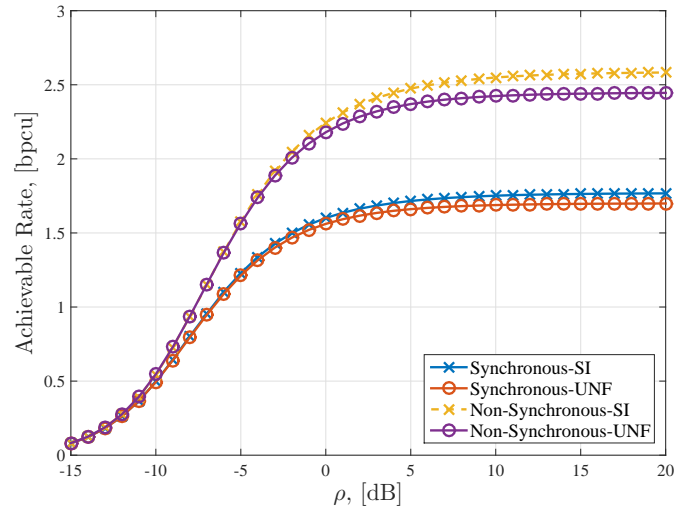


Figure 4.3: Comparison of the average achievable rates in (4.27) and (4.30) for $M = 100$, $\sigma_\phi^2 = \sigma_\theta^2 = 10^{-3}$ and $\tau_d = 500$ as a function of ρ , [dB].

Chapter 5

Contributions of the Dissertation

This dissertation focuses on the performance of Massive MIMO in frequency-selective channels and in the presence of phase noise impairments. The achievable rate performance of a frequency-selective Massive MIMO downlink is studied and compared with multi-carrier transmission. The achievable rate performance of a frequency-selective Massive MIMO uplink is studied in the presence of phase noise and estimated channel impulse responses is also analyzed. The study is extended to a frequency-flat phase-noise-impaired Massive MIMO uplink with zero-forcing receivers. The maximum-likelihood detector in a frequency-flat phase-noise-impaired SIMO channel is studied in the presence of training-based estimated channel impulse responses. The achievable rate performance of a phase-noise-impaired Massive MIMO-OFDM uplink is also studied.

5.1 Included Papers

Brief summaries of the papers included in this dissertation are as follows:

Paper A: On the Optimality of Single-Carrier Transmission in Large-Scale Antenna Systems

Authored by Antonios Pitarokoilis, Saif Khan Mohammed, and Erik G. Larsson.

Published in the IEEE Wireless Communications Letters, 2012 [71].

A single carrier transmission scheme is presented for the frequency selective multi-user (MU) multiple-input single-output (MISO) Gaussian Broadcast Channel (GBC) with a base station (BS) having M antennas and K single antenna users. The proposed transmission scheme has low complexity and for $M \gg K$ it is shown to achieve near optimal sum-rate performance at low transmit power to receiver noise power ratio. Additionally, the proposed transmission scheme results in an equalization-free receiver and does not require any MU resource allocation and associated control signaling overhead. Also, the sum-rate achieved by the proposed transmission scheme is shown to be independent of the channel power delay profile (PDP). In terms of power efficiency, the proposed transmission scheme also exhibits an $O(M)$ array power gain. Simulations are used to confirm analytical observations.

Paper B: Uplink Performance of Time-Reversal MRC in Massive MIMO Systems Subject to Phase Noise

Authored by Antonios Pitarokoilis, Saif Khan Mohammed, and Erik G. Larsson.

Published in the IEEE Transactions on Wireless Communications, 2015 [72]. This work is an extension of the conference paper [73].

Multi-user multiple-input multiple-output (MU-MIMO) cellular systems with an excess of base station (BS) antennas (Massive MIMO) offer unprecedented multiplexing gains and radiated energy efficiency. Oscillator phase noise is introduced in the transmitter and receiver radio frequency chains and severely degrades the performance of communication systems. We study the effect of oscillator phase noise in frequency-selective Massive MIMO systems with imperfect channel state information (CSI). In particular, we consider two distinct operation modes, namely when the phase

noise processes at the M BS antennas are identical (synchronous operation) and when they are independent (non-synchronous operation). We analyze a linear and low-complexity time-reversal maximum-ratio combining (TR-MRC) reception strategy. For both operation modes we derive a lower bound on the sum-capacity and we compare their performance. Based on the derived achievable sum-rates, we show that with the proposed receive processing an $O(\sqrt{M})$ array gain is achievable. Due to the phase noise drift the estimated effective channel becomes progressively outdated. Therefore, phase noise effectively limits the length of the interval used for data transmission and the number of scheduled users. The derived achievable rates provide insights into the optimum choice of the data interval length and the number of scheduled users.

Paper C: Achievable Rates of ZF Receivers in Massive MIMO with Phase Noise Impairments

Authored by Antonios Pitarokoilis, Saif Khan Mohammed, and Erik G. Larsson.

Published in the proceedings of the Asilomar Conference on Signals, Systems and Computers, 2013 [74].

The effect of oscillator phase noise on the sum-rate performance of large multi-user multiple-input multiple-output (MU-MIMO) systems is studied. A Rayleigh fading MU-MIMO uplink channel is considered, where channel state information (CSI) is acquired via training. The base station (BS), which is equipped with an excess of antenna elements, M , uses the channel estimate to perform zero-forcing (ZF) detection. A lower bound on the sum-rate performance is derived. It is shown that the proposed receiver structure exhibits an $O(\sqrt{M})$ array power gain. Additionally, the proposed receiver is compared with earlier studies that employ maximum ratio combining and it is shown that it can provide significant sum-rate performance gains at the medium and high signal-to-noise-ratio (SNR) regime. Further, the expression of the achievable sum rate provides new insights on the effect of various parameters on the overall system performance.

Paper D: ML Detection in Phase Noise Impaired SIMO Channels with Uplink Training

Authored by Antonios Pitarokoilis, Emil Björnson, and Erik G. Larsson.

Published in the IEEE Transactions on Communications, 2016 [75]. This is an extension of the conference paper [76].

The problem of maximum likelihood (ML) detection in training-assisted single-input multiple-output (SIMO) systems with phase noise impairments is studied for two different scenarios, i.e. the case when the channel is deterministic and known (constant channel) and the case when the channel is stochastic and unknown (fading channel). Further, two different operations with respect to the phase noise sources are considered, namely, the case of identical phase noise sources and the case of independent phase noise sources over the antennas. In all scenarios the optimal detector is derived for a very general parameterization of the phase noise distribution. Further, a high signal-to-noise-ratio (SNR) analysis is performed to show that symbol-error-rate (SER) floors appear in all cases. The SER floor in the case of identical phase noise sources (for both constant and fading channels) is independent of the number of antenna elements. In contrast, the SER floor in the case of independent phase noise sources is reduced when increasing the number of antenna elements (for both constant and fading channels). Finally, the system model is extended to multiple data channel uses and it is shown that the conclusions are valid for these setups, as well.

Paper E: Performance of the Massive MIMO Uplink with OFDM and Phase Noise

Authored by Antonios Pitarokoilis, Emil Björnson, and Erik G. Larsson.

Submitted to the IEEE Communications Letters.

The performance of multi-user Massive MIMO-OFDM uplink systems in the presence of base station (BS) phase noise impairments is investigated. Closed-form achievable rate expressions are rigorously derived under two different operations, namely the case of a common oscillator (synchronous operation) at the BS and the case of independent oscillators at each BS antenna (non-synchronous operation). It is observed that the non-synchronous operation exhibits superior performance due to the averaging of intercarrier interference. Further, radiated power scaling laws are derived, which are identical to the phase-noise-free case.

5.2 Not Included Papers

The following publications by the author are not included in the dissertation either because they do not fit within the main scope of the dissertation, or they were earlier versions of the journal publications included in the dissertation.

- **A. Pitarokoilis**, S. K. Mohammed, and E. G. Larsson, 'Effect of oscillator phase noise on uplink performance of large MU-MIMO systems', in *Proc. 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2012, Monticello, IL, USA, pp. 1190–1197, Oct. 2012
- **A. Pitarokoilis**, E. Björnson, and E. G. Larsson, 'Optimal Detection in Training Assisted SIMO Systems with Phase Noise Impairments', in *IEEE International Conference on Communications (ICC)*, London, U.K., pp. 2597–2602, June 2015
- E. Björnson, M. Matthaiou, **A. Pitarokoilis**, and E. G. Larsson, 'Distributed Massive MIMO in Cellular Networks: Impact of Imperfect Hardware & Number of Oscillators', in *23rd European Signal Processing Conference (EUSIPCO)*, 2015, Nice, France, pp. 2436–2440, Aug. 31 2015–Sept. 4 2015

5.3 Future Research Directions

In this dissertation various aspects of Massive MIMO systems with wide-band transmission and phase noise impairments are investigated. However, there are still many problems that are worthwhile to be investigated. In addition, the observations made here give rise to new research questions. In this section, we attempt to identify problems that are related with the present dissertation and could be considered by researchers in the future.

- In Paper A it was shown that single-carrier downlink transmission with perfect CSI at the BS can be close to optimal in the low spectral efficiency regime. The work could be extended to the imperfect CSI

case and to the uplink transmission. In multi-cell frequency-flat Massive MIMO systems pilot contamination appears due to finite length of the coherence interval. It is reasonable to ask whether pilot contamination persists in frequency-selective channels. Can the frequency selectivity of the channel be used in order to suppress pilot contamination from users in neighboring cells?

- In Paper B and Paper C achievable sum-rates were derived for frequency-selective and frequency-flat uplink Massive MIMO systems with phase noise impairments. It is worthwhile to extend the results to the frequency-selective downlink and to other linear schemes, such as regularized zero forcing (RZF) precoding in the downlink or MMSE reception in the uplink. Towards this research direction, prior work includes [77], where the authors use asymptotic random matrix theory and provide capacity bounds for the case of frequency-flat downlink transmission with ZF precoding.
- In Paper B and Paper C the CSI was acquired via a simple training schedule, i.e., the training interval is at the beginning of the coherence interval and is not interleaved with the uplink data interval. Even though, the qualitative behavior of the achievable rates is not expected to change with some different scheduling of pilots, e.g., pilots at the end of the coherence interval or pilots interleaved with data transmission, it is useful to investigate whether there is an optimal pilot scheduling. Towards this direction, the authors in [69] have provided initial results for the flat-fading case and for more general modeling of hardware impairments that includes phase noise and additive hardware impairments.
- The derived achievable rates in Paper B and Paper C assumed that no phase noise estimation and compensation scheme was employed. However, it is reasonable to assume that there might be some phase noise tracking mechanism, such as the one proposed in [78]. The investigation of the gains in terms of achievable rate with phase noise compensation is a very important research topic. In [79] a simulation-based approach for the calculation of achievable rates is provided for a SISO AWGN phase noise channel. The extension of this work to Massive MIMO is particularly challenging and highly non-trivial, as the system model must be augmented with the effects of imperfect CSI, multi-user interference and channel fading.
- In Paper D the optimal, in the ML sense, detector with training based

CSI was derived for a basic system model. It is important to investigate whether similar conclusions can be drawn in the downlink or if there is any fundamental difference. Also, in prior work, suboptimal algorithms based on factor graphs [80,81] that perform well for single-user SISO [82] and MIMO [83] have been investigated. However, the provided algorithms appear to be too complex for Massive MIMO systems. Nevertheless, the works in [82], [83] and Paper D can be used as a starting point for novel, low complexity detection algorithms that are tailored to multi-user Massive MIMO setups with estimated CSI.

- In Paper E achievable rates were derived on Massive MIMO-OFDM uplink systems with phase noise impairments when perfect knowledge of the propagation channel is available at the receiver. The extension of the study to estimated CSI is a very relevant research problem. In the presence of phase noise the orthogonality of subchannels is destroyed and intercarrier interference appears. In addition, in OFDM systems often certain subcarriers in the same OFDM symbol are reserved for pilots and the rest for data transmission. Due to the presence of intercarrier interference, the estimated CSI is correlated with the transmitted data. This phenomenon complicates the analytical treatment of Massive MIMO with estimated CSI and phase noise.
- The Wiener phase noise model was used in the present dissertation. However, this model is an accurate description of free-running LOs with uncorrelated circuit noise sources. The analysis could be extended to more complex phase noise models that are tailored to LOs with certain spectral characteristics. Finally, recently the symbol-spaced phase noise model is being revisited [33] and phase noise has been treated in the continuous time [84]. Also, in [34, 85, 86] multi-sample receivers for the Wiener phase noise channel have been studied and it was shown that an increase in the pre-log factor can be achieved. The extension of such receivers to the Massive MIMO setup could open new relevant research problems.
- The implementation of testbeds that could help verify the results presented in this dissertation is an important topic of study. Further, the outcome of these studies can help the researchers in communications to refine the applied models and make more accurate predictions of the true performance.

Bibliography

- [1] Cisco, "White paper: Cisco visual networking index: Global mobile data traffic forecast update, 2015-2020," tech. rep., February 2016.
- [2] Ericsson, "White paper: 5G systems enabling industry and society transformation," Tech. Rep. Uen 284 23-3244, January 2015.
- [3] E. Telatar, "Capacity of multi-antenna gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.
- [4] G. Foschini and M. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, Mar. 1998.
- [5] D. Gesbert, M. Kountouris, R. W. Heath Jr., C.-B. Chae, and T. Sälzer, "Shifting the MIMO paradigm," *IEEE Signal Processing Magazine*, vol. 24, pp. 36–46, Sept. 2007.
- [6] J. Winters, "Optimum combining for indoor radio systems with multiple users," *IEEE Transactions on Communications*, vol. 35, pp. 1222–1230, Nov. 1987.
- [7] S. C. Swales, M. A. Beach, D. J. Edwards, and J. P. McGeehan, "The performance enhancement of multibeam adaptive base-station antennas for cellular land mobile radio systems," *IEEE Transactions on Vehicular Technology*, vol. 39, pp. 56–67, Feb. 1990.
- [8] S. Anderson, M. Millnert, M. Viberg, and B. Wahlberg, "An adaptive array for mobile communication systems," *IEEE Transactions on Vehicular Technology*, vol. 40, pp. 230–236, Feb. 1991.

- [9] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, pp. 3590–3600, Nov. 2010.
- [10] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Transactions on Communications*, vol. 61, pp. 1436–1449, Apr. 2013.
- [11] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?," *IEEE Journal on Selected Areas in Communications*, vol. 31, pp. 160–171, Feb. 2013.
- [12] G. Gonzalez, *Foundations of Oscillator Circuit Design*. Norwood, MA, USA: Artech House, Inc., 2006.
- [13] A. Hajimiri and T. Lee, "A general theory of phase noise in electrical oscillators," *IEEE Journal of Solid-State Circuits*, vol. 33, pp. 179–194, Feb. 1998.
- [14] A. Demir, A. Mehrotra, and J. Roychowdhury, "Phase noise in oscillators: a unifying theory and numerical methods for characterization," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 47, pp. 655–674, May 2000.
- [15] J. G. Proakis, *Digital Communications*. McGraw-Hill, 1995.
- [16] G. Kaplan and U. Ram, "Bounds on performance for the noisy reference PSK channel," *IEEE Transactions on Communications*, vol. 38, pp. 1699–1707, Oct. 1990.
- [17] P. Y. Kam, S. Teo, Y. K. Some, and T. T. Tjhung, "Approximate results for the bit error probability of binary phase shift keying with noisy phase reference," *IEEE Transactions on Communications*, vol. 41, pp. 1020–1022, July 1993.
- [18] B. Razavi, "A study of phase noise in CMOS oscillators," *IEEE Journal of Solid-State Circuits*, vol. 31, pp. 331–343, Mar. 1996.
- [19] T. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*. Cambridge University Press, 2004.
- [20] J. B. Johnson, "Thermal agitation of electricity in conductors," *Physical Review*, vol. 32, pp. 97–109, July 1928.

- [21] H. Nyquist, "Thermal agitation of electric charge in conductors," *Physical Review*, vol. 32, pp. 110–113, July 1928.
- [22] W. Schottky, "Über spontane Stromschwankungen in verschiedenen Elektrizitätsleitern," *Annalen der Physik*, vol. 362, no. 23, pp. 541–567, 1918.
- [23] M. S. Keshner, "1/f noise," *Proceedings of the IEEE*, vol. 70, pp. 212–218, Mar. 1982.
- [24] B. Mandelbrot, "Some noises with 1/f spectrum, a bridge between direct current and white noise," *IEEE Transactions on Information Theory*, vol. 13, pp. 289–298, Apr. 1967.
- [25] D. Leeson, "A simple model of feedback oscillator noise spectrum," *Proceedings of the IEEE*, vol. 54, pp. 329–330, Feb. 1966.
- [26] T. Lee and A. Hajimiri, "Oscillator phase noise: a tutorial," *IEEE Journal of Solid-State Circuits*, vol. 35, pp. 326–336, Mar. 2000.
- [27] J. H. Dauwels, *On graphical models for communications and machine learning: algorithms, bounds, and analog implementation*. PhD thesis, ETH Zürich, Hartung-Gorre Verlag, Konstanz, 2006. Diss., Eidgenössische Technische Hochschule ETH Zürich, Nr. 16365, 2005.
- [28] D. Petrovic, W. Rave, and G. Fettweis, "Effects of phase noise on OFDM systems with and without PLL: Characterization and compensation," *IEEE Transactions on Communications*, vol. 55, pp. 1607–1616, Aug. 2007.
- [29] P. Billingsley, *Probability and Measure, Anniversary Edition*. Wiley, 2012.
- [30] A. Demir, "Phase noise in oscillators: DAEs and colored noise sources," in *IEEE/ACM International Conference on Computer-Aided Design, 1998. ICCAD 98. Digest of Technical Papers. 1998*, pp. 170–177, Nov. 1998.
- [31] H. L. van Trees, *Detection, Estimation, and Modulation Theory, Part I*. New York: Wiley Interscience, 2001.
- [32] G. Colavolpe, "Communications over phase-noise channels: A tutorial review," in *6th Advanced Satellite Multimedia Systems Conference (ASMS) and 12th Signal Processing for Space Communications Workshop (SPSC), 2012*, pp. 316–327, Sept. 2012.

- [33] M. Martalò, C. Tripodi, and R. Raheli, "On the information rate of phase noise-limited communications," in *Information Theory and Applications Workshop (ITA), 2013*, pp. 1–7, Feb. 2013.
- [34] H. Ghozlan and G. Kramer, "Models and information rates for Wiener phase noise channels," *CoRR*, vol. abs/1503.03130, 2015.
- [35] T. Pollet, M. Van Bladel, and M. Moeneclaey, "BER sensitivity of OFDM systems to carrier frequency offset and wiener phase noise," *IEEE Transactions on Communications*, vol. 43, pp. 191–193, Feb/Mar/Apr 1995.
- [36] L. Tomba, "On the effect of Wiener phase noise in OFDM systems," *IEEE Transactions on Communications*, vol. 46, pp. 580–583, May 1998.
- [37] Q. Zou, A. Tarighat, and A. Sayed, "Compensation of phase noise in OFDM wireless systems," *IEEE Transactions on Signal Processing*, vol. 55, pp. 5407–5424, Nov. 2007.
- [38] T. Schenk, *RF Imperfections in High-rate Wireless Systems: Impact and Digital Compensation*. Springer, 2008.
- [39] P. Mathecken, T. Riihonen, S. Werner, and R. Wichman, "Performance analysis of OFDM with wiener phase noise and frequency selective fading channel," *IEEE Transactions on Communications*, vol. 59, pp. 1321–1331, May 2011.
- [40] P. Mathecken, T. Riihonen, N. Tchamov, S. Werner, M. Valkama, and R. Wichman, "Characterization of OFDM radio link under PLL-based oscillator phase noise and multipath fading channel," *IEEE Transactions on Communications*, vol. 60, pp. 1479–1485, June 2012.
- [41] A. Viterbi, *Principles of Coherent Communication*. McGraw-Hill Inc., 1966.
- [42] G. J. Foschini, R. D. Gitlin, and S. B. Weinstein, "On the selection of a two-dimensional signal constellation in the presence of phase jitter and Gaussian noise," *Bell System Technical Journal*, vol. 52, pp. 927–965, July-Aug. 1973.
- [43] A. Lapidoth, "On phase noise channels at high SNR," in *Proc. of the 2002 IEEE Information Theory Workshop*, pp. 1–4, Oct. 2002.

- [44] B. Goebel, R.-J. Essiambre, G. Kramer, P. J. Winzer, and N. Hanik, "Calculation of mutual information for partially coherent Gaussian channels with applications to fiber optics," *IEEE Transactions on Information Theory*, vol. 57, pp. 5720–5736, Sept. 2011.
- [45] G. Durisi, A. Tarable, C. Camarda, R. Devassy, and G. Montorsi, "Capacity bounds for MIMO microwave backhaul links affected by phase noise," *IEEE Transactions on Communications*, vol. 62, pp. 920–929, Mar. 2014.
- [46] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, UK: Cambridge Univ. Press, 2004.
- [47] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. New York, NY, USA: Cambridge University Press, 2012.
- [48] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE Journal on Selected Areas in Communications*, vol. 21, pp. 684–702, June 2003.
- [49] P. Viswanath, D. N. C. Tse, and V. Anantharam, "Asymptotically optimal water-filling in vector multiple-access channels," *IEEE Transactions on Information Theory*, vol. 47, pp. 241–267, Jan 2001.
- [50] L. Li and A. J. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels .I. ergodic capacity," *IEEE Transactions on Information Theory*, vol. 47, pp. 1083–1102, Mar. 2001.
- [51] L. Li and A. J. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels .II. outage capacity," *IEEE Transactions on Information Theory*, vol. 47, pp. 1103–1127, Mar. 2001.
- [52] G. Caire and S. Shamai, "On the achievable throughput of a multi-antenna gaussian broadcast channel," *IEEE Transactions on Information Theory*, vol. 49, pp. 1691–1706, July 2003.
- [53] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector gaussian broadcast channel and uplink-downlink duality," *IEEE Transactions on Information Theory*, vol. 49, pp. 1912–1921, Aug. 2003.
- [54] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of gaussian MIMO broadcast channels," *IEEE Transactions on Information Theory*, vol. 49, pp. 2658–2668, Oct. 2003.

- [55] W. Yu and J. M. Cioffi, "Sum capacity of gaussian vector broadcast channels," *IEEE Transactions on Information Theory*, vol. 50, pp. 1875–1892, Sept. 2004.
- [56] N. Jindal, S. Vishwanath, and A. Goldsmith, "On the duality of gaussian multiple-access and broadcast channels," *IEEE Transactions on Information Theory*, vol. 50, pp. 768–783, May 2004.
- [57] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the gaussian multiple-input multiple-output broadcast channel," *IEEE Transactions on Information Theory*, vol. 52, pp. 3936–3964, Sept. 2006.
- [58] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, pp. 40–60, Jan. 2013.
- [59] E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, pp. 186–195, Feb. 2014.
- [60] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Massive MU-MIMO downlink TDD systems with linear precoding and downlink pilots," in *Communication, Control, and Computing (Allerton), 2013 51st Annual Allerton Conference on*, pp. 293–298, Oct 2013.
- [61] H. Q. Ngo, *Massive MIMO: Fundamentals and System Designs*. Linköping Studies in Science and Technology. Dissertations No. 1642, Linköping University, Sweden, 2015.
- [62] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Transactions on Information Theory*, vol. 46, pp. 933–946, May 2000.
- [63] A. Lapidoth and S. Shamai, "Fading channels: how perfect need "perfect side information" be?," *IEEE Transactions on Information Theory*, vol. 48, pp. 1118–1134, May 2002.
- [64] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Transactions on Information Theory*, vol. 49, pp. 951 – 963, Apr. 2003.
- [65] T. L. Marzetta, "How much training is required for multiuser MIMO?," in *Proc. 40th Asilomar Conference on Signals, Systems and Computers*, pp. 359 –363, Nov. 2006.

- [66] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Transactions on Wireless Communications*, vol. 10, pp. 2640–2651, Aug. 2011.
- [67] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Transactions on Information Theory*, vol. 60, pp. 7112–7139, Nov. 2014.
- [68] H. Yang and T. L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, pp. 172–179, Feb. 2013.
- [69] E. Björnson, M. Matthaiou, and M. Debbah, "Massive MIMO with non-ideal arbitrary arrays: Hardware scaling laws and circuit-aware design," *IEEE Transactions on Wireless Communications*, vol. 14, pp. 4353–4368, Aug. 2015.
- [70] E. Björnson, M. Matthaiou, A. Pitarokoilis, and E. G. Larsson, "Distributed Massive MIMO in cellular networks: Impact of imperfect hardware and number of oscillators," in *23rd European Signal Processing Conference (EUSIPCO), 2015*, pp. 2436–2440, Aug. 2015.
- [71] A. Pitarokoilis, S. K. Mohammed, and E. G. Larsson, "On the optimality of single-carrier transmission in large-scale antenna systems," *IEEE Wireless Communications Letters*, vol. 1, pp. 276–279, Aug. 2012.
- [72] A. Pitarokoilis, S. K. Mohammed, and E. G. Larsson, "Uplink performance of time-reversal MRC in massive MIMO systems subject to phase noise," *IEEE Transactions on Wireless Communications*, vol. 14, pp. 711–723, Feb. 2015.
- [73] A. Pitarokoilis, S. K. Mohammed, and E. G. Larsson, "Effect of oscillator phase noise on the uplink performance of large MU-MIMO systems," in *Proc. of the 50th Allerton Conference on Communication Control and Computing*, pp. 1190–1197, Oct. 2012.
- [74] A. Pitarokoilis, S. K. Mohammed, and E. G. Larsson, "Achievable rates of ZF receivers in massive MIMO with phase noise impairments," in *Proc. of the 47th Asilomar Conference on Signals, Systems and Computers, 2013*, pp. 1004–1008, Nov. 2013.

- [75] A. Pitarokoilis, E. Björnson, and E. G. Larsson, "ML detection in phase noise impaired SIMO channels with uplink training," *IEEE Transactions on Communications*, vol. 64, pp. 223–235, Jan. 2016.
- [76] A. Pitarokoilis, E. Björnson, and E. G. Larsson, "Optimal detection in training assisted SIMO systems with phase noise impairments," in *IEEE International Conference on Communications (ICC), 2015*, pp. 2597–2602, June 2015.
- [77] R. Krishnan, M. Khanzadi, N. Krishnan, Y. Wu, A. Graell i Amat, T. Eriksson, and R. Schober, "Linear Massive MIMO precoders in the presence of phase noise – a large-scale analysis," *IEEE Transactions on Vehicular Technology*, vol. PP, no. 99, pp. 1–1, 2015.
- [78] H. Mehrpouyan, A. A. Nasir, S. D. Blostein, T. Eriksson, G. K. Karagiannidis, and T. Svensson, "Joint estimation of channel and oscillator phase noise in MIMO systems," *IEEE Transactions on Signal Processing*, vol. 60, pp. 4790–4807, Sept. 2012.
- [79] L. Barletta, M. Magarini, S. Pecorino, and A. Spalvieri, "Upper and lower bounds to the information rate transferred through first-order markov channels with free-running continuous state," *IEEE Transactions on Information Theory*, vol. 60, pp. 3834–3844, July 2014.
- [80] F. R. Kschischang, B. J. Frey, and H. A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on Information Theory*, vol. 47, pp. 498–519, Feb 2001.
- [81] H. A. Loeliger, "An introduction to factor graphs," *IEEE Signal Processing Magazine*, vol. 21, pp. 28–41, Jan 2004.
- [82] G. Colavolpe, A. Barbieri, and G. Caire, "Algorithms for iterative decoding in the presence of strong phase noise," *IEEE Journal on Selected Areas in Communications*, vol. 23, pp. 1748–1757, Sept. 2005.
- [83] R. Krishnan, G. Colavolpe, A. Graell i Amat, and T. Eriksson, "Algorithms for joint phase estimation and decoding for MIMO systems in the presence of phase noise and quasi-static fading channels," *IEEE Transactions on Signal Processing*, vol. 63, pp. 3360–3375, July 2015.
- [84] L. Barletta and G. Kramer, "On continuous-time white phase noise channels," in *2014 IEEE International Symposium on Information Theory*, pp. 2426–2429, June 2014.

-
- [85] H. Ghozlan and G. Kramer, "Multi-sample receivers increase information rates for Wiener phase noise channels," in *2013 IEEE Global Communications Conference (GLOBECOM)*, pp. 1897–1902, Dec 2013.
- [86] H. Ghozlan and G. Kramer, "Phase modulation for discrete-time Wiener phase noise channels with oversampling at high SNR," in *2014 IEEE International Symposium on Information Theory*, pp. 1554–1557, June 2014.

Part II

Included Papers

PAPER A

On the Optimality of Single-Carrier Transmission in Large-Scale Antenna Systems

Refereed article published in the IEEE Wireless
Communications Letters 2012.

©2012 IEEE. The layout has been revised and the
bibliography has been updated.

On the Optimality of Single-Carrier Transmission in Large-Scale Antenna Systems

Antonios Pitarokoilis, Saif Khan Mohammed, and Erik G. Larsson

Abstract

A single carrier transmission scheme is presented for the frequency selective multi-user (MU) multiple-input single-output (MISO) Gaussian Broadcast Channel (GBC) with a base station (BS) having M antennas and K single antenna users. The proposed transmission scheme has low complexity and for $M \gg K$ it is shown to achieve near optimal sum-rate performance at low transmit power to receiver noise power ratio. Additionally, the proposed transmission scheme results in an equalization-free receiver and does not require any MU resource allocation and associated control signaling overhead. Also, the sum-rate achieved by the proposed transmission scheme is shown to be independent of the channel power delay profile (PDP). In terms of power efficiency, the proposed transmission scheme also exhibits an $O(M)$ array power gain. Simulations are used to confirm analytical observations.

1 Introduction

Multiple-input multiple-output (MIMO) systems have attracted significant research interest during the last decade due to various advantages they promise, both in single user [1] and multiuser channels [2]. It has been recently shown that the employment of an excess of antennas at the BS (very large MIMO) offers unprecedented array and multiplexing gains both in the uplink and in the downlink [3, 4]. The array gain offered by very large MIMO systems allows for power savings that scale as $1/M$ and $1/\sqrt{M}$ with perfect and imperfect channel state information (CSI) respectively, where M is the number of BS antennas [5]. The multiplexing gains offered by very large MIMO allows tens of users to be allocated the entire system bandwidth simultaneously. This eliminates to a large extent the need for resource allocation and the associated control signaling overhead. Since each user communicates over the whole system bandwidth, even low per user spectral efficiencies can result in very high per user throughput. In a MU-MISO GBC with K users and $M \gg K$ (very large MIMO), a low per user spectral efficiency implies an operating regime where the ratio of the total transmit power to the receiver additive noise power is small. Since MU interference at each receiver is proportional to the total transmit power, the additive noise dominates over MU interference and therefore even suboptimal precoding algorithms (like beamforming with the conjugate transpose of the channel gain matrix) have near optimal performance.

Previous results for very large MIMO systems have only considered frequency flat channels [3–5]. In this paper we consider a MU-MISO *frequency selective* GBC with $M \gg K$. For this channel OFDM (OFDMA) is an attractive transmission scheme as it facilitates scheduling in the frequency domain and simplifies receiver equalization. However, there is a substantial price to pay for this. OFDM comes at a loss in spectral and power efficiency owing to the insertion of cyclic prefix. Moreover, the signals resulting from OFDM modulation have a very large peak-to-average ratio, requiring the RF power amplifiers to work with a large power backoff and in an operating regime where they have low efficiency. For this reason, single-carrier or single-carrier-like modulation schemes like DFT-precoded OFDM are often used when there are stringent requirements on power efficiency of the RF amplifiers. Single-carrier signals have a much lower peak-to-average ratio and can be shaped to have constant envelope even in multiuser MIMO systems [6].

The contributions made in this paper are summarized as follows. 1) We firstly propose a low complexity single carrier transmission scheme for the frequency selective MU-MISO GBC. 2) At low total transmit power to receiver noise ratio, the proposed transmission scheme is shown to effectively suppress intersymbol interference (ISI) and MU interference at each receiver, thereby achieving near optimal sum-rate performance. 3) Additionally, the proposed scheme does not require any receiver equalization. Also, its simplicity allows for separate, decentralized computation at each BS antenna. 4) An achievable information sum-rate is derived for the proposed scheme. This sum-rate is further shown to be invariant of the channel PDP. 5) In terms of power efficiency, the proposed scheme is shown to exhibit an array power gain proportional to the number of BS antennas.

2 System Model

A frequency selective MU-MISO downlink channel is considered, with M BS antennas and K single antenna users. The channel between the m -th transmit antenna and the k -th user is modeled as a finite impulse response (FIR) filter with L taps. The l -th channel tap is given by $\sqrt{d_l[k]}h_l^*[m, k]$, where $h_l^*[m, k]$ and $d_l[k]$ model the fast and slow varying components, respectively. In this paper we assume a model where $h_l^*[m, k]$ is fixed during the transmission of a block of N symbols and varies independently from one block to another. However, the slowly varying component (i.e. $\sqrt{d_l[k]}$) is assumed to be fixed throughout the entire communication. We further assume $h_l^*[m, k]$ to be i.i.d. $\mathcal{CN}(0, 1)$ distributed. $d_l[k] \geq 0$, $l = 0, \dots, L - 1$ models the PDP of the frequency selective channel for the k -th user.¹ Let $x_m[i]$ be the symbol transmitted from transmit antenna m at time i . The received signal at user k at time i is then given by

$$y_k[i] = \sum_{l=0}^{L-1} \sum_{m=1}^M \sqrt{d_l[k]} h_l^*[m, k] x_m[i-l] + n_k[i], \quad (1)$$

where $n_k[i]$ is the $\mathcal{CN}(0, 1)$ distributed AWGN at the k -th receiver at time i . Define $\mathbf{y}[i] \triangleq [y_1[i], \dots, y_K[i]]^T \in \mathbb{C}^K$ to be the vector of received user symbols at time i . Similarly, let $\mathbf{x}[i] \triangleq [x_1[i], \dots, x_M[i]]^T \in \mathbb{C}^M$ be

¹PDP determines the distribution of the received power across different channel taps.

the transmitted vector at time i . Let $\mathbf{n}[i] \triangleq [n_1[i], \dots, n_K[i]]^T$, with independent components. The received signal vector at time i is given by $\mathbf{y}[i] = \sum_{l=0}^{L-1} \mathbf{D}_l^{1/2} \mathbf{H}_l^H \mathbf{x}[i-l] + \mathbf{n}[i]$, where $\mathbf{D}_l \triangleq \text{diag}\{d_l[1], \dots, d_l[K]\}$, and $\mathbf{H}_l \in \mathbb{C}^{M \times K}$ is a matrix whose (m, k) -th element is $h_l[m, k]$. Also the channel PDP for each user is normalized such that

$$\sum_{l=0}^{L-1} d_l[k] = 1, \quad \forall k = 1, \dots, K. \quad (2)$$

The BS is assumed to have full CSI, whereas the users have knowledge of the channel statistics only.²

Let $s_k[i]$ denote the information symbol to be communicated to the k -th user at time i . The information symbol vector $\mathbf{s}[i] = [s_1[i], \dots, s_K[i]]^T$ is considered to have i.i.d. $\mathcal{CN}(0, 1)$ components, i.e. $\mathbb{E}[\mathbf{s}[i] \mathbf{s}^H[i+j]] = \mathbf{I}_K \delta_j$, $\mathbb{E}[\mathbf{s}[i] \mathbf{s}^T[i+j]] = \mathbf{0}$. In this paper we propose a precoding scheme, where the transmitted vector at time i is given by

$$\mathbf{x}[i] = \sqrt{\frac{\rho_f}{MK}} \sum_{l=0}^{L-1} \mathbf{H}_l \mathbf{D}_l^{1/2} \mathbf{s}[i+l], \quad (3)$$

where $\rho_f \triangleq \mathbb{E}[\|\mathbf{x}[i]\|^2]$ is the long-term average total power radiated by the BS antennas. In the following, we derive an achievable sum-rate for the proposed precoder in (3).

3 Achievable Sum-Rate

The bounding technique of [7,8] is used here to obtain an achievable rate. In the following, a set of achievable rates is presented. For notational brevity we define $\mathbf{v}_l[k] \triangleq \mathbf{H}_l \mathbf{D}_l^{1/2} \mathbf{e}_k$, where \mathbf{e}_k is the all-zero vector except for the k -th component which is equal to 1. Using (1) and (3) the signal received by user k at time i is given by

²In a time division duplex (TDD) system, CSI at the BS can be acquired through uplink training and exploiting the uplink-downlink channel reciprocity.

$$y_k[i] = \underbrace{\left(\sqrt{\frac{\rho_f}{MK}} \sum_{l=0}^{L-1} \mathbb{E} [\mathbf{v}_l^H[k] \mathbf{v}_l[k]] \right)}_{\text{Desired Signal Term}} s_k[i] + \underbrace{n'_k[i]}_{\text{Effective Noise Term}}, \quad (4)$$

where³

$$\begin{aligned} n'_k[i] \triangleq & \underbrace{\sqrt{\frac{\rho_f}{MK}} \left(\sum_{l=0}^{L-1} \mathbf{v}_l^H[k] \mathbf{v}_l[k] - \sum_{l=0}^{L-1} \mathbb{E} [\mathbf{v}_l^H[k] \mathbf{v}_l[k]] \right)}_{\text{Additional Interference Term (IF)}} s_k[i] \\ & + \underbrace{\sqrt{\frac{\rho_f}{MK}} \sum_{\substack{a=1-L \\ a \neq 0}}^{L-1} \sum_{l=\max(a,0)}^{\min(L-1+a, L-1)} \mathbf{v}_l^H[k] \mathbf{v}_{l-a}[k] s_k[i-a]}_{\text{Intersymbol Interference (ISI)}} \\ & + \underbrace{\sqrt{\frac{\rho_f}{MK}} \sum_{\substack{q=1 \\ q \neq k}}^K \sum_{a=1-L}^{L-1} \sum_{l=\max(a,0)}^{\min(L-1+a, L-1)} \mathbf{v}_l^H[k] \mathbf{v}_{l-a}[q] s_q[i-a]}_{\text{Multiuser Interference (MUI)}} \\ & + \underbrace{n_k[i]}_{\text{AWGN}} \end{aligned} \quad (5)$$

is the effective noise term. This term includes (i) the IF term which represents the variation of the desired signal around its mean, (ii) the ISI term between the current symbol of user k , i.e. $s_k[i]$, and the symbols intended to the same user at other time instances (i.e. $s_k[i+j]$, $j \neq 0$), (iii) the MUI term due to the information symbols intended for other users and, (iv) the AWGN term. In the proposed precoder, each user's codeword is long enough such that it spans across multiple coherence intervals. With long codewords, the effective variance of $n'_k[i]$ is no longer dependent on a particular channel realization but only depends on the channel statistics. From this it follows that the desired signal $s_k[i]$ is uncorrelated with the

³Following [7, 8], we have split the coefficient of the term $\sqrt{\frac{\rho_f}{MK}} \sum_{l=0}^{L-1} \mathbf{v}_l[k]^H \mathbf{v}_l[k] s_k[i]$ into a sum of its mean value (which is known to the receiver) and the deviation around its mean.

effective noise $n'_k[i]$, i.e. $\mathbb{E}[s_k[i]n'_k[i]] = 0$, where the expectation is taken over the channel realizations, the information symbols and additive noise. Therefore, with long codewords the channel is effectively an additive noise channel with the noise $n'_k[i]$ being non-Gaussian and uncorrelated to the information symbol $s_k[i]$. Further, the user has perfect knowledge of its channel statistic and therefore it knows the scaling factor $\sum_{l=0}^{L-1} \mathbb{E}[\mathbf{v}_l^H[k]\mathbf{v}_l[k]]$. Hence, an achievable information rate for the channel in (4) is given by considering the worst case uncorrelated additive noise having the same variance as $n'_k[i]$. Given that the data signal $\mathbf{s}[i]$ is Gaussian, the worst uncorrelated additive noise is circularly symmetric Gaussian distributed with the same variance as $n'_k[i]$. Therefore, the following information rate is achievable for the k -th user

$$R_k = \log_2(1 + S_k/\text{Var}(n'_k[i])) \quad (6)$$

where $S_k = \mathbb{E}_{s_k[i]} \left[\left| \sqrt{\frac{\rho_f}{MK}} \sum_{l=0}^{L-1} \mathbb{E}[\mathbf{v}_l^H[k]\mathbf{v}_l[k]] s_k[i] \right|^2 \right]$ is the average power of the desired signal term in (4) and $\text{Var}(n'_k[i]) \triangleq \mathbb{E}[|n'_k[i] - \mathbb{E}[n'_k[i]]|^2]$.

Proposition 1. *The variance of $n'_k[i]$ is invariant of any PDP that satisfies (2), and is given by*

$$\text{Var}(n'_k[i]) = \rho_f + 1. \quad (7)$$

Proof. Using (5), the effective noise variance is given by

$$\begin{aligned} \text{Var}(n'_k[i]) &= \frac{\rho_f}{K} \sum_{q=1}^K \sum_{a=1}^{L-1} \sum_{l=a}^{L-1} (d_{l-a}[k]d_l[q] + d_l[k]d_{l-a}[q]) \\ &\quad + \frac{\rho_f}{K} \sum_{q=1}^K \sum_{l=0}^{L-1} d_l[k]d_l[q] + 1, \end{aligned} \quad (8)$$

where the expectation is taken over the statistics of \mathbf{H}_l , $l = \{0, \dots, L-1\}$, $\mathbf{s}[i+a]$, $a = \{1-L, \dots, L-1\}$ and $n_k[i]$. Define $\mathbf{\Delta} \in \mathbb{R}^{K \times L}$ such that $[\mathbf{\Delta}]_{i,j} = d_{j-1}[i]$ and let $\mathbf{1} \in \{1\}^{L \times L}$ denote the matrix with all entries equal

to one. Then, (8) can be expressed as

$$\begin{aligned}
\text{Var}(n'_k[i]) &= \frac{\rho_f}{K} \sum_{q=1}^K \sum_{a=1}^{L-1} \sum_{l=a}^{L-1} e_k^T \Delta (e_{l-a+1} e_{l+1}^T \\
&\quad + e_{l+1} e_{l-a+1}^T) \Delta^T e_q + \frac{\rho_f}{K} \sum_{q=1}^K e_k^T \Delta \Delta^T e_q + 1 \\
&= \frac{\rho_f}{K} \sum_{q=1}^K e_k^T \Delta \mathbf{1} \Delta^T e_q + 1. \tag{9}
\end{aligned}$$

From (2) it follows that $e_k^T \Delta \mathbf{1} = [1 \dots 1]$. Using this fact in (9) completes the proof. \square

It is apparent from (7) that the variance of the effective noise consists of the variance of the white noise term (which is 1) and the variance of the sum of interference terms (which is ρ_f). In the following we provide an explanation as to why the variance of the effective noise term is invariant of the PDP. Note that the precoder in (3) is like a matching pre-filter whose impulse response is a time reversed and complex-conjugated image of the channel impulse response (CIR). Due to this special structure of the proposed precoder, $n'_k[i]$ is composed of terms which consist of all non-zero auto-correlation lags of the CIR for the k -th user (ISI term in (5)), as well as all cross-correlation lags between the CIR of user k and the CIR of the remaining $(K - 1)$ users (MUI term in (5)). The effective MUI in $y_k[i]$ from the symbols intended for the q -th user, depends only upon the total power in all channel correlation lags between the CIR's of the k -th and the q -th user. Due to the same channel and information symbol statistics for all users, the effective MUI in $y_k[i]$ from each of the remaining $(K - 1)$ users is identical, and is independent of the individual PDPs (the total power in the cross-correlation lags depends only upon the total power in the CIR for each user, which is independent of k due to (2)).

Further, the useful signal term in $y_k[i]$ is proportional to the zero-lag auto-correlation of the CIR for the k -th user. This zero-lag auto-correlation (i.e. maximum gain combining of the lags) is proportional to the total channel power gain (combining all taps) from the M BS antennas to the k -th user and is therefore $O(M)$. The average power of the desired signal term in (4) is given by $\mathbb{E}_{s_k[i]} \left[\left| \sqrt{\frac{\rho_f}{MK}} \sum_{l=0}^{L-1} \mathbb{E} [\mathbf{v}_l^H[k] \mathbf{v}_l[k]] s_k[i] \right|^2 \right] = \rho_f M / K$. Using

this fact and (7) in (6), the achievable rate R_k for user k is given by $R_k = \log_2(1 + \rho_f M / (K\rho_f + K))$. The achievable sum-rate is therefore given by

$$R_{\text{sum}}(\rho_f, M, K) = \sum_{k=1}^K R_k = K \log_2 \left(1 + \frac{\rho_f M}{K\rho_f + K} \right). \quad (10)$$

For the sake of comparison, we also consider a co-operative upper bound on the sum-capacity of the frequency selective GBC.⁴ Essentially, we get an upper bound by considering the users to be co-operative, which reduces the MU channel to a single user MIMO channel, with perfect CSI at both the transmitter and the receiver. We further consider transmission in time with large blocks (block size $\gg L$), where in each block the last few transmit vectors are zeros so as to avoid any inter-block interference. The sum-capacity for this single user MIMO block channel is given by beamforming along the right singular vectors of the effective channel matrix, thus transforming the channel into a set of parallel channels. Gaussian symbols are communicated over the parallel channels and power allocation is given by the waterfilling scheme.

With i.i.d complex normal entries in \mathbf{H}_l , it is clear that for fixed K , $\frac{\mathbf{H}_l^H \mathbf{H}_l}{M} \rightarrow I_K$ as $M \rightarrow \infty$. Therefore for $M \gg K$, the K singular values of \mathbf{H}_l are all roughly equal to \sqrt{M} (i.e., the power gain for each parallel channel is $\approx M$). With a uniform power allocation of ρ_f/K across parallel channels, the co-operative upper bound on the ergodic sum-capacity of the GBC is given by

$$C_{\text{coop}}(\rho_f, M, K) \approx K \log_2(1 + \rho_f M / K). \quad (11)$$

We conclude our analysis with two propositions on the near-optimality and the array gain of the proposed precoder.

Proposition 2. *When $\rho_f \ll 1$ and $M \gg K$, $R_{\text{sum}} \approx C_{\text{coop}}$ and the proposed precoder is near-optimal.*

Proof. Observe that when $\rho_f \ll 1$, the effective noise variance, $\text{Var}(n'_k) = \rho_f + 1 \approx 1$ (essentially the additive white noise dominates over the interference terms in (5)). It follows that, $K\rho_f + K \approx K$ and therefore the expressions in (10) and (11) are approximately equal. \square

⁴The sum-capacity of the MIMO GBC is known. However, for the results reported in this paper, it suffices to consider only the co-operative upper bound on the sum-capacity.

Proposition 3. *The proposed precoder exhibits an $O(M)$ array power gain.*

Proof. For the proposed precoder, using (10) the minimum transmit power ρ_f required to achieve a fixed desired sum-rate R_{sum} with K users and M BS antennas is given by $\rho_f(M) = \frac{K(2^{R_{\text{sum}}/K} - 1)}{M + K(2^{R_{\text{sum}}/K} - 1)}$. Since $\lim_{M \rightarrow \infty} \frac{1}{M} \frac{\rho_f(1)}{\rho_f(M)} = \frac{1}{1 + K(2^{R_{\text{sum}}/K} - 1)} > 0$ from [9] it follows that the proposed precoder achieves an $O(M)$ array power gain. \square

This implies that for a sufficiently large M , $\rho_f(M)/\rho_f(1) \propto 1/M$ (i.e. the total transmitted power can be reduced linearly by increasing the number of BS antennas). A similar analysis of the co-operative sum-capacity (see (11)) reveals that the array power gain achieved by a sum-capacity achieving scheme is also $O(M)$.

4 Simulation Results

In the following, representative simulation results are presented, where the performance of proposed precoder is compared to the co-operative upper bound. Throughout the conducted simulations, the PDP is exponential with $L = 4$ and $d_l[k] = \frac{e^{-\theta_k l}}{\sum_{i=0}^3 e^{-\theta_k i}}$, $l = \{0, \dots, 3\}$, where $\theta_k = (k - 1)/5$, $k = \{1, \dots, K\}$. As proved in Proposition 1, the achievable sum-rate is invariant of the PDP. Hence, any other PDP which satisfies (2), would also yield the same results. Proposition 2 is supported by Fig. 1, where the sum-rate is plotted as a function of ρ_f , for $M = 50$ and $K = 10$. The sum-rate performance of the proposed precoder is given both by the theoretical expression (10) and via simulations. Similarly, the co-operative sum-capacity upper bound is calculated via simulations and by the approximation in (11). For $\rho_f \ll 1$ (0 dB), as can be seen in Fig. 1, the performance of the proposed precoder is similar to the upper bound, implying optimality. Note that as ρ_f increases, the interference terms dominate over the white noise term in (5) and the effective noise variance is therefore $\rho_f + 1 \approx \rho_f$. Hence, as $\rho_f \rightarrow \infty$ the sum-rate of the proposed precoder saturates to the value $K \log_2(1 + M/K) = 25.85$ bpcu. It can also be seen that the approximation to the sum-capacity upper bound is tight.

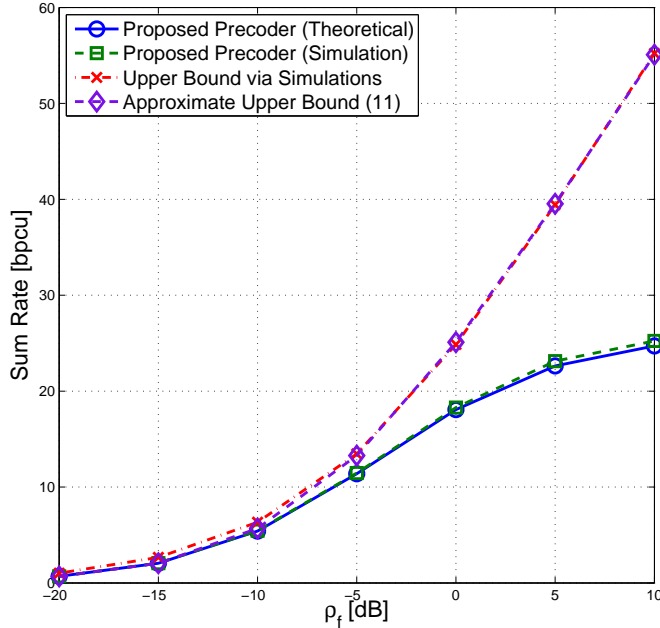


Figure 1: Sum rate of the proposed precoder and the co-operative sum-capacity upper bound vs ρ_f , calculated for $K = 10$ users, $M = 50$ BS antennas.

The analytical result in Proposition 3 is supported by Fig. 2, where for a fixed number of users and a fixed per user rate of 1 bpcu, the minimum total transmit power required is plotted as a function of the number of BS antennas. In Fig. 2 it is observed that the minimum transmit power required by the proposed precoder can be reduced by roughly 3dB with every doubling in the number of the BS antennas (for sufficiently large M). This implies the achievability of an $O(M)$ array power gain, as stated in Proposition 3. In Fig. 2 it is also observed that for sufficiently large values of M the total transmit power required by the proposed precoder is roughly equal to the total transmit power required by a sum-capacity achieving scheme. Further, for the sake of comparison, consider a typical scenario, where OFDM is used. Let ρ_f^{OFDM} denote the total transmit power for OFDM transmission. Under OFDM transmission with $M \gg K$, it can be shown that the per user ergodic information rate (in i.i.d. Rayleigh fading channel) is given by

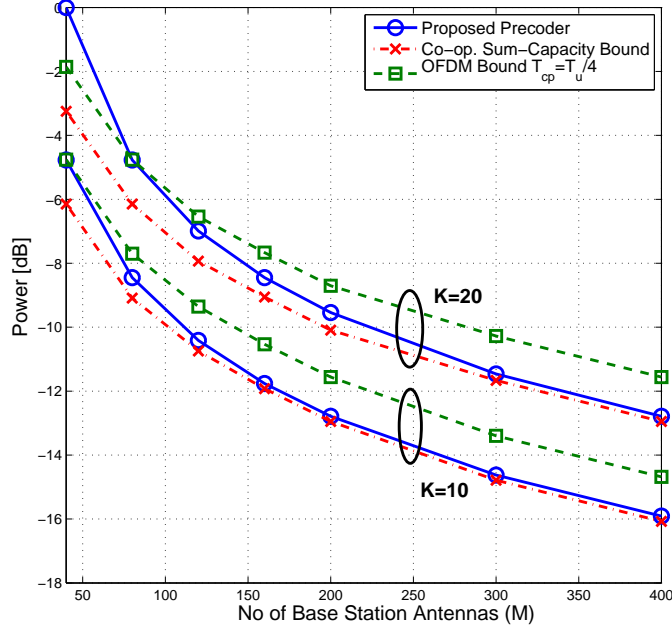


Figure 2: Minimum required transmit power to achieve a fixed per user information rate $r = 1$ bpcu as a function of the number of BS antennas.

$$r \approx \frac{T_u}{T_u + T_{cp}} \log_2 \left(1 + \rho_f^{\text{OFDM}} \frac{M}{K} \right), \quad (12)$$

where T_{cp} is the duration of the cyclic prefix and T_u is the duration of the useful signal.⁵ From (12) it follows that, to achieve an ergodic per user information rate of r bpcu the minimum required total transmit power is given by $\rho_f^{\text{OFDM}}(r) \approx \frac{K}{M} (2^{r(T_u + T_{cp})/T_u} - 1)$.

⁵Note that in practice, modern wireless standards employ $X > 1$ OFDM symbols per coherence time interval. Each OFDM symbol consists of N_u channel uses for data transmission and N_{cp} channel uses for the cyclic prefix. This means that XN_{cp} channel uses in each coherence time interval are used for non-data transmission. In contrast, in the proposed precoding scheme only N_{cp} channel uses per coherence time interval are used for non-data transmission. (These N_{cp} channel uses are used for zero-padding at the beginning of each coherence interval.) Note that in practical wireless standards $X \gtrsim 10$, which implies that the proposed precoder makes better use of available channel bandwidth.

For a given desired per user ergodic information rate, the additional total transmit power required under OFDM transmission when compared to an optimal GBC sum-capacity achieving scheme is upper bounded by $\rho_f^{\text{OFDM}}(r)/\rho_f^{\text{coop}}(r)$, where $\rho_f^{\text{coop}}(r) = (2^r - 1)K/M$ is roughly equal to the required transmit power for the co-operative sum-capacity bound to be rK bpcu (see (11)). The additional transmit power required under OFDM transmission is therefore given by $\frac{2^{r(1+T_{\text{cp}}/T_u)} - 1}{2^r - 1}$. Since $\frac{2^{r(1+T_{\text{cp}}/T_u)} - 1}{2^r - 1} > 1$ and the total transmit power required by the proposed precoder is roughly equal to that required by a sum-capacity achieving scheme ($M \gg K$ and $\rho_f \ll 1$ (see Proposition 2)), it can be concluded that the proposed precoder is more power efficient than OFDM transmission for large M/K . As an example, for a typical IEEE 802.11a scenario with $T_{\text{cp}} = T_u/4$, a desired per user information rate $r = 1$ bpcu and $M \gg K$, this additional transmit power required under OFDM transmission when compared to the proposed precoder is 1.39 dB. The minimum transmit power required under OFDM transmission is also plotted in Fig. 2, where it can be seen that for $M > 4K$ the proposed precoding scheme is more power efficient than OFDM transmission and requires no equalization at the receiver (Note that equalization in OFDM receivers requires FFT processing).

Bibliography

- [1] G. Foschini and M. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, Mar. 1998.
- [2] D. Gesbert, M. Kountouris, R. W. Heath Jr., C.-B. Chae, and T. Sälzer, "Shifting the MIMO paradigm," *IEEE Signal Processing Magazine*, vol. 24, pp. 36–46, Sept. 2007.
- [3] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, pp. 40–60, Jan. 2013.
- [4] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, pp. 3590–3600, Nov. 2010.
- [5] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Transactions on Communications*, vol. 61, pp. 1436–1449, Apr. 2013.
- [6] S. K. Mohammed and E. G. Larsson, "Per-antenna constant envelope precoding for large multi-user MIMO systems," in *IEEE Transactions on Communications*, vol. 61, no. 3, pp. 1059–1071, Mar. 2013.
- [7] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Transactions on Information Theory*, vol. 49, pp. 951–963, Apr. 2003.
- [8] T. L. Marzetta, "How much training is required for multiuser MIMO?," in *Proc. 40th Asilomar Conference on Signals, Systems and Computers*, pp. 359–363, Nov. 2006.

- [9] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. Cambridge, UK: Cambridge Univ. Press, 2004.

PAPER B

Uplink Performance of Time-Reversal MRC in Massive MIMO Systems Subject to Phase Noise

Refereed article published in the IEEE Transactions on
Wireless Communications 2015.

©2015 IEEE. The layout has been revised, minor typographical errors
have been corrected and the bibliography has been updated.

Uplink Performance of Time-Reversal MRC in Massive MIMO Systems Subject to Phase Noise

Antonios Pitarokoilis, Saif Khan Mohammed, and Erik G. Larsson

Abstract

Multi-user multiple-input multiple-output (MU-MIMO) cellular systems with an excess of base station (BS) antennas (Massive MIMO) offer unprecedented multiplexing gains and radiated energy efficiency. Oscillator phase noise is introduced in the transmitter and receiver radio frequency chains and severely degrades the performance of communication systems. We study the effect of oscillator phase noise in frequency-selective Massive MIMO systems with imperfect channel state information (CSI). In particular, we consider two distinct operation modes, namely when the phase noise processes at the M BS antennas are identical (synchronous operation) and when they are independent (non-synchronous operation). We analyze a linear and low-complexity time-reversal maximum-ratio combining (TR-MRC) reception strategy. For both operation modes we derive a lower bound on the sum-capacity and we compare their performance. Based on the derived achievable sum-rates, we show that with the proposed receive processing an $O(\sqrt{M})$ array gain is achievable. Due to the phase noise drift the estimated effective channel becomes progressively outdated. Therefore, phase noise effectively limits the length of the interval used for data transmission and the number of scheduled users. The derived achievable rates provide insights into the optimum choice of the data interval length and the number of scheduled users.

1 Introduction

Multiple-input multiple-output (MIMO) technology offers substantial performance gains in wireless links [1]. The spatial degrees of freedom enable many users to share the same time-frequency resources, paving the way for multi-user MIMO (MU-MIMO) systems [2]. MU-MIMO systems with an excess of BS antennas, termed as Massive MIMO or large-scale MIMO, have recently attracted significant interest [3–5]. They promise a significant increase in the total cell throughput by means of simple signal processing. At the same time, the radiated power can be scaled down with the number of BS antennas, M , while maintaining a desired sum-rate. More specifically, in [6] the authors show that in a MU-MIMO uplink with linear receivers and imperfect channel state information (CSI), by increasing the number of BS antennas from 1 to M , one can reduce the total transmit power by a factor $O(\sqrt{M})$ while maintaining a fixed per-user information rate. In [7] the authors report an improved result for channels with arbitrary channel covariance matrices. The crucial assumption in Massive MIMO is that the squared Euclidean norm of the channel vector of each user grows as $O(M)$, whereas the inner products between channel vectors of different users grow at a lesser rate. This assumption can be justified in the MU-MIMO setting since the users are typically separated by many wavelengths, which implies that their channel vectors become asymptotically (in the number of BS antennas) orthogonal. Extensive measurements have confirmed the validity of this assumption [4,5].

Phase noise is inevitable in communication systems due to imperfections in the circuitry of the local oscillators that are used for the conversion of the baseband signal to passband and vice versa. To be specific, phase noise is the instantaneous drift of the phase of the carrier wave and results in a widening of the power spectral density of the generated waveform. Phase noise causes a partial loss of coherency between the channel estimate and the true channel gain during data transmission. This can result in severe degradation of the system performance.

In MIMO an array power gain is obtained by coherently combining signals received by several antennas, using estimated channel responses. Since phase noise distorts the received data, it is crucial to examine its effect on the performance. Significant research work is available on phase noise. However, most of it is concerned with single-user single-antenna

multi-carrier transmission, since multi-carrier transmission is more sensitive to phase noise compared to single-carrier transmission [8]. In [9] a method to calculate the bit-error-rate (BER) of a single-user orthogonal frequency division multiplexing (OFDM) system impaired with phase noise is provided. Reference [10] studies the signal-to-interference-and-noise-ratio (SINR) degradation in OFDM and proposes a method to mitigate the effect of phase noise. In [11] a method to characterize phase noise in OFDM systems is developed and an algorithm to compensate for the degradation is described. Finally, in [12] the authors propose a method to jointly estimate the channel coefficients and the phase noise in a single-user MIMO system and an associated phase noise mitigation algorithm.

From an information-theoretic point of view, the calculation of capacity of phase noise channels is challenging. To the best of our knowledge, the exact capacity of typical phase noise-impaired channels under realistic models is not known. The behavior of the capacity of such channels is only known asymptotically for some cases in the high signal-to-noise-ratio (SNR) regime [13]. In [14] the authors derive a non-asymptotic upper bound on the capacity of a single-user deterministic MIMO channel impaired with Wiener phase noise, which is tight in the high-SNR regime. In [15], the authors consider the performance of Massive MIMO systems with hardware impairments. Their model is suitable for the residual hardware impairments after the application of appropriate compensation algorithms.

To the authors' knowledge, we present the first analysis of the effect of Wiener phase noise in a multi-user multi-antenna scenario with imperfect channel state information where single-carrier transmission is used. Specifically, we consider a single-cell frequency-selective MU-MIMO uplink, where a number of non-cooperative users transmit independent data streams to a base station having a large number of antennas. Since the channel is assumed to be unknown, CSI is acquired via uplink training. There are phase noise sources both at the transmitters and at the receiver. In this paper we extend the work presented in [16]. We consider and compare two distinct cases. In the first case, which is termed *synchronous operation* mode, the phase noise processes at the BS antennas are identical. In the second case, which is termed *non-synchronous operation* mode, the phase noise processes at the BS antennas are independent. These two operation modes correspond to the cases of a common phase reference versus independent phase references, respectively. A time-reversal maximum-ratio combining

(TR-MRC) strategy is proposed and achievable sum-rates are derived for both operation modes.

Based on the derived expressions of the achievable sum-rates, we show that for a fixed desired per-user information rate, by doubling the number of BS antennas, the total transmit power can be reduced by a factor of $\sqrt{2}$. This is the same scaling law as without phase noise [6]. We observe that the use of independent phase noise sources can yield higher sum-rate performance and we support this interesting result by a simple toy example for which the exact capacity is calculated. Furthermore, the achievable rate expressions reveal a fundamental trade-off between the length of the time interval spent on data transmission and the sum-rate performance. The rate expressions also provide valuable insight into the optimum number of scheduled users.

2 System Model

We consider a frequency-selective MU-MIMO uplink channel with M BS antennas and K single-antenna users. The channel between the k -th user and the m -th BS antenna is modeled as a finite impulse response (FIR) filter with L symbol-spaced channel taps. The l -th channel tap is given by $g_{m,k,l} \triangleq \sqrt{d_{k,l}}h_{m,k,l}$, where $h_{m,k,l}$ and $d_{k,l}$ model the fast and slow time-varying components, respectively. We assume a block fading model where $h_{m,k,l}$ is fixed during the transmission of a block of $N_c \triangleq N_D + (K + 3)L - 3$ symbols and varies independently from one block to another. N_D denotes the number of channel uses utilized for data transmission (see Fig. 1). We further assume that the channel fading process is ergodic. The parameters $d_{k,l} \geq 0$, $l = 0, \dots, L - 1$ model the power delay profile (PDP) of the frequency-selective channel for the k -th user. Since $\{d_{k,l}\}$ vary slowly with time and spatial location, we assume them to be fixed for the entire communication and independent of m . We further assume $h_{m,k,l}$ to be independent and identically distributed (i.i.d.) zero-mean and unit-variance proper complex random variables. The i.i.d. assumption is justified in [4, 5, 17].¹ Further, the PDP for every user is normalized such that the average received

¹We note that with the i.i.d. assumption on the channel gains, the captured energy increases linearly with the number of BS antennas, M . This is not reasonable if M grows unbounded. However, this deficiency of the model takes effect only for exorbitantly large values of M which do not lie in the regime of our interest [5], [4], [7].

power is independent of the length of the channel impulse response, L . Therefore, it holds that

$$\sum_{l=0}^{L-1} \mathbb{E} \left[|\sqrt{d_{k,l}} h_{m,k,l}|^2 \right] = \sum_{l=0}^{L-1} d_{k,l} = \alpha_k, \quad (1)$$

for $1 \leq k \leq K$. The positive constants, α_k , account for different propagation losses between users and are assumed to be fixed throughout the communication. The BS is assumed to have perfect knowledge of all the PDPs. Finally, we assume exact knowledge of the channel statistics at the BS, but not of the particular channel realizations.

2.1 Phase Noise Model

Phase noise is introduced at the transmitter during up-conversion, when the baseband signal is multiplied with the carrier generated by the local oscillator. The phase of the generated carrier drifts randomly, resulting in a phase distortion of the transmitted signal. A similar phenomenon also happens at the receiver side during down-conversion of the bandpass signal to baseband. In the following, θ_k , $k = 1, \dots, K$ denotes the phase noise process at the k -th single-antenna user. Since the users have different local oscillators, the transmitter phase noise processes are assumed to be mutually independent. On the other hand, at the receiver side two distinct operation modes are considered. We term these operation modes as *synchronous* and *non-synchronous* operation depending on whether the phase noise processes at the BS antennas are identical or independent. For the synchronous case, all BS antennas are subject to the same phase noise process and ϕ denotes this common phase noise process at each BS antenna. This models the scenario of a centralized BS with a single oscillator feeding the down-conversion module in each receiver. For the case of non-synchronous operation, ϕ_m , $m = 1, \dots, M$ denotes the phase noise process at the m -th BS antenna. This models a completely distributed scenario where each BS antenna uses a distinct oscillator for down-conversion. We further assume that the phase noise processes θ_k , $k = 1, \dots, K$ and ϕ (or ϕ_m , $m = 1, \dots, M$) for the case of synchronous (or non-synchronous) operation mode are mutually independent.

In this study each phase noise process is modeled as an independent Wiener process, which is a well-established model [11, 18]. Therefore, the discrete-

time phase noise process at the k -th user at time i is given by²

$$\theta_k[i] = \theta_k[i-1] + w_k^t[i], \quad (2)$$

where $w_k^t[i] \sim \mathcal{N}(0, \sigma_\theta^2)$ are independent identically distributed zero-mean Gaussian increments with variance $\sigma_\theta^2 \triangleq 4\pi^2 f_c^2 c_\theta T_s$, f_c is the carrier frequency, T_s is the symbol interval and c_θ is a constant that depends on the oscillator. Depending on the operation mode, the phase noise processes $\phi[i]$ and $\phi_m[i]$ at the M BS antennas are defined in a manner similar to (2), where the increments have variance $\sigma_\phi^2 \triangleq 4\pi^2 f_c^2 c_\phi T_s$.

2.2 Received Signal

Let $x_k[i]$ be the symbol transmitted from the k -th user at time i . The received sample at the m -th BS antenna element at time i is then given by, for the non-synchronous operation

$$y_m[i] = \sqrt{P} \sum_{k=1}^K \sum_{l=0}^{L-1} e^{-j\phi_m[i]} g_{m,k,l} e^{j\theta_k[i-l]} x_k[i-l] + n_m[i], \quad (3)$$

where $n_m[i] \sim \mathcal{CN}(0, \sigma^2)$ represents noise at the m -th receiver at time i , which is distributed as circularly symmetric complex Gaussian.³ Each user transmits a stream of i.i.d. $\mathcal{CN}(0, 1)$ information symbols (i.e., $x_k[i] \sim \mathcal{CN}(0, 1)$), that are independent of the information symbols of the other users. P denotes the average uplink transmitted power from each user.

3 Transmission Scheme and Receive Processing

We consider a block-based uplink transmission scheme. A transmission block of N_c channel uses consists of KL channel uses dedicated to uplink channel training followed by a preamble of $L - 1$ channel uses, where i.i.d.

²The discrete-time phase noise model is used since we will be working with the discrete-time complex baseband representation of the transmit and receive signals.

³In the following we will present only the expressions of the non-synchronous mode. The expressions for the synchronous operation are obtained easily by substituting $\phi_1[i] \equiv \dots \equiv \phi_M[i] \equiv \phi[i]$. In Sections 4–6, when the expressions of the two distinct modes differ in a non-obvious way, both expressions will be given explicitly.

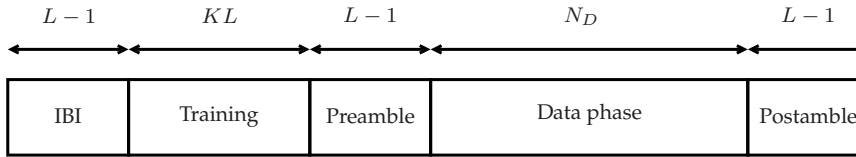


Figure 1: The transmission block is assumed to span a coherence interval, $N_c \triangleq N_D + (K + 3)L - 3$. In each block, the first KL channel uses (cu) are utilized for pilot based channel estimation and N_D cu are utilized for data transmission. An all-zero block, a preamble and a postamble of $L - 1$ cu each are added due to the edge effects of the channel.

$\mathcal{CN}(0, 1)$ non-information symbols are sent. The data interval of N_D channel uses comes after that and a postamble of $L - 1$ channel uses is appended at the end of the coherence interval, where i.i.d. $\mathcal{CN}(0, 1)$ non-information symbols are sent. The inclusion of the preamble and postamble accounts for the edge effects introduced due to the intersymbol interference. This way the subsequent analysis is valid for all the N_D channel uses during data transmission and no separate analysis for the edges of the data interval is required. At the beginning of each coherence interval an all-zero block of $L - 1$ channel uses is prepended to eliminate inter-block interference (IBI) (see Fig. 1).

3.1 Channel Estimation

For coherent demodulation, the BS needs to estimate the uplink channel. This is facilitated through the transmission of uplink pilot symbols during the training phase of each transmission block.⁴ The users transmit uplink training signals sequentially in time, i.e., at any given time only one user is transmitting uplink training signals and all other users are silent. To be precise, the k -th user sends an impulse of amplitude $\sqrt{P_p KL}$ at the $(k - 1)L$ -th channel use and is idle for the remaining portion of the training phase. Here, P_p is the average power transmitted by a user during the training phase. We choose the proposed training sequence since it allows for a very simple channel estimation scheme at the BS and since it facilitates our

⁴In this paper we deal only with uplink transmission. In Massive MIMO Time Division Duplex (TDD) operation pilots are transmitted on the uplink. The number of required pilots scales with the number of terminals, K , but not the number of BS antennas, M , making Massive MIMO scalable with respect to M [3], [4].

derivation of achievable rates. However, many of our results, such as partial loss of coherency due to Wiener phase noise and monotonic decrease in performance with increased variance of the phase noise increments, are expected to be qualitatively valid also for other (but not necessarily all possible) training schemes. Therefore, using (3), the signal received at the m -th BS receiver at time $(k-1)L+l$, $l=0, \dots, L-1$, $k=1, \dots, K$ is given by, for non-synchronous operation

$$y_m[(k-1)L+l] = \sqrt{P_p K L} g_{m,k,l} e^{j(\theta_k[(k-1)L] - \phi_m[(k-1)L+l])} + n_m[(k-1)L+l]. \quad (4)$$

Based on (4), we derive the maximum likelihood (ML) estimate of the effective channel $g_{m,k,l} e^{j(\theta_k[(k-1)L] - \phi_m[(k-1)L+l])}$. The corresponding channel estimates are then given by, for non-synchronous operation

$$\begin{aligned} \hat{g}_{m,k,l} &= \frac{1}{\sqrt{P_p K L}} y_m[(k-1)L+l] \\ &= g_{m,k,l} e^{-j\phi_m[(k-1)L+l]} e^{j\theta_k[(k-1)L]} + \frac{1}{\sqrt{P_p K L}} n_m[(k-1)L+l]. \end{aligned} \quad (5)$$

We observe that the channel estimate is distorted by the AWGN and by the phase noise of the local oscillators at the user and at the BS.

3.2 Time-Reversal Maximum Ratio Combining (TR-MRC)

Using (3), the received signal during the data phase is given by, for non-synchronous operation

$$y_m[i] = \sqrt{P_D} \sum_{k=1}^K \sum_{l=0}^{L-1} e^{-j\phi_m[i]} g_{m,k,l} e^{j\theta_k[i-l]} x_k[i-l] + n_m[i], \quad (6)$$

where $i \in \mathcal{I}_d$, $\mathcal{I}_d \triangleq \{(K+1)L-1, \dots, (K+1)L+N_D-2\}$ and P_D is the per-user average transmit power constraint during the data phase. Motivated by the need for low-complexity detection, we consider the TR-MRC receiver at the BS. The TR-MRC receiver convolves the received symbols, $y_m[i]$, with the complex conjugate of the time-reversed estimated channel impulse response. The detected symbol, $\hat{x}_k[i]$, is given by

$$\hat{x}_k[i] = \sum_{l=0}^{L-1} \sum_{m=1}^M \hat{g}_{m,k,l}^* y_m[i+l], \quad (7)$$

where $(\cdot)^*$ denotes the complex conjugation operation.

4 Achievable Sum-Rate

We use the information sum-rate as the performance metric for quantifying the effects of phase noise. To this end, using (5) and (6) for the non-synchronous operation, (7) is written as

$$\hat{x}_k[i] = A_k[i]x_k[i] + \text{ISI}_k[i] + \text{MUI}_k[i] + \text{AN}_k[i], \quad (8)$$

where it holds for the non-synchronous operation that

$$A_k[i] \triangleq \sqrt{P_D} \sum_{m=1}^M \sum_{l=0}^{L-1} |g_{m,k,l}|^2 \vartheta \binom{m,k,k}{i,l,l} \quad (9)$$

$$\text{ISI}_k[i] \triangleq \sqrt{P_D} \sum_{m=1}^M \sum_{l=0}^{L-1} \sum_{\substack{p=0 \\ p \neq l}}^{L-1} g_{m,k,l}^* g_{m,k,p} \vartheta \binom{m,k,k}{i,l,p} x_k[i+l-p] \quad (10)$$

$$\text{MUI}_k[i] \triangleq \sqrt{P_D} \sum_{m=1}^M \sum_{\substack{q=1 \\ q \neq k}}^K \sum_{l=0}^{L-1} \sum_{p=0}^{L-1} g_{m,k,l}^* g_{m,q,p} \vartheta \binom{m,k,q}{i,l,p} x_q[i+l-p] \quad (11)$$

$$\begin{aligned} \text{AN}_k[i] \triangleq & \sqrt{\frac{P_D}{P_p K L}} \sum_{m=1}^M \sum_{q=1}^K \sum_{l=0}^{L-1} \sum_{p=0}^{L-1} g_{m,q,p} \times \\ & e^{-j(\phi_m[i+l] - \theta_q[i+l-p])} n_m[(k-1)L+l] x_q[i+l-p] \\ & + \sum_{m=1}^M \sum_{l=0}^{L-1} \hat{g}_{m,k,l}^* n_m[i+l], \end{aligned} \quad (12)$$

where $\vartheta \binom{m,k,q}{i,l,p} \triangleq e^{j(\theta_q[i+l-p] - \theta_k[(k-1)L] - \phi_m[i+l] + \phi_m[(k-1)L+l])}$. In (8), $A_k[i]x_k[i]$ is the desired signal term for the k -th user, $\text{ISI}_k[i]$ stands for the intersymbol interference for user k at time i , caused by the information symbols of the k -th user transmitted at other time instances, $\text{MUI}_k[i]$ denotes the multi-user interference due to the information symbols of the other users and finally $\text{AN}_k[i]$ is an aggregate noise term that incorporates the effects of the channel estimation error and the receiver AWGN noise, $n_m[i]$. The expressions for the terms in (8) for the synchronous operation are obtained from (9)-(12) by substituting $\phi_1[i] \equiv \dots \equiv \phi_M[i] \equiv \phi[i]$.

In the following, we derive an achievable information rate for the k -th user. Similar capacity bounding techniques have been used earlier in e.g. [19, 20]. In (8), we add and subtract the term $\mathbb{E}[A_k[i]]x_k[i]$, where the expectation is taken over the channel gains, $g_{m,k,l}$, and the phase noise processes, θ_k, ϕ for the synchronous operation and θ_k, ϕ_m for the non-synchronous operation. We relegate the variation around this term, i.e., $\text{IF}_k[i] \triangleq (A_k[i] - \mathbb{E}[A_k[i]])x_k[i]$, to an effective noise term. This results in the following equivalent expression

$$\hat{x}_k[i] = \mathbb{E}[A_k[i]]x_k[i] + \text{EN}_k[i], \quad (13)$$

where

$$\text{EN}_k[i] \triangleq \text{IF}_k[i] + \text{ISI}_k[i] + \text{MUI}_k[i] + \text{AN}_k[i], \quad (14)$$

is the effective additive noise term. In (13) the detected symbol, $\hat{x}_k[i]$, is a sum of two *uncorrelated* terms (i.e., $\mathbb{E}[(\mathbb{E}[A_k[i]]x_k[i])(\text{EN}_k[i])^*] = 0$). The importance of the equivalent representation in (13) is that the scaling factor $\mathbb{E}[A_k[i]]$ of the desired information symbol is a constant, which is known at the BS since the BS has knowledge of the channel statistics. The exact probability distribution of $\text{EN}_k[i]$ is difficult to compute. However, its variance can be easily calculated given that the channel statistics is known at the BS. Therefore, (13) describes an effective single-user single-input single-output (SISO) additive noise channel, where the noise is zero mean, has known variance and is uncorrelated with the desired signal term. From the expressions for $A_k[i]$ and $\text{EN}_k[i]$ in (9) and (14), the mean value of $A_k[i]$ and the variance of $\text{EN}_k[i]$ is given by two propositions that follow.

Proposition 1. *The mean value of $A_k[i]$ in both operation modes is given by*

$$\mathbb{E}[A_k[i]] = \sqrt{P_D} M \alpha_k e^{-\frac{\sigma_\phi^2 + \sigma_\theta^2}{2}(i - (k-1)L)}. \quad (15)$$

Proof. We prove the statement for the non-synchronous operation. The

proof for the synchronous operation is nearly identical. From (9), we have

$$\begin{aligned}
\mathbb{E}[A_k[i]] &= \mathbb{E} \left[\sqrt{P_D} \sum_{m=1}^M \sum_{l=0}^{L-1} |g_{m,k,l}|^2 \vartheta \begin{pmatrix} m, k, k \\ i, l, l \end{pmatrix} \right] \\
&\stackrel{(a)}{=} \sqrt{P_D} \mathbb{E} \left[e^{-j(\theta_k[(k-1)L] - \theta_k[i])} \sum_{m=1}^M \sum_{l=0}^{L-1} \mathbb{E} \left[|g_{m,k,l}|^2 \right] \right. \\
&\quad \cdot \mathbb{E} \left[e^{-j(\phi_m[i+l] - \phi_m[(k-1)L+l])} \right] \left. \right] \\
&\stackrel{(b)}{=} \sqrt{P_D} e^{-\frac{\sigma_\theta^2}{2}(i-(k-1)L)} \sum_{m=1}^M \sum_{l=0}^{L-1} d_{k,l} e^{-\frac{\sigma_\phi^2}{2}(i-(k-1)L)} \\
&\stackrel{(c)}{=} \sqrt{P_D} M \alpha_k e^{-\frac{\sigma_\phi^2 + \sigma_\theta^2}{2}(i-(k-1)L)}.
\end{aligned}$$

In (a) we have used the fact that the channel realizations, $g_{m,k,l}$, the phase noise at the BS, ϕ_m , and the phase noise at the k -th user, θ_k , are mutually independent random processes. The equality (b) is a consequence of the Wiener phase noise model. That is, after a time interval, $\Delta t = i - (k-1)L$, the phase drift of an oscillator is a zero mean Gaussian random variable with variance that is proportional to Δt ,

$$\begin{aligned}
U_{\phi_m} &\triangleq \phi_m[i+l] - \phi_m[(k-1)L+l] \sim \mathcal{N}(0, \sigma_\phi^2(i-(k-1)L)), \\
U_{\theta_k} &\triangleq \theta_k[i] - \theta_k[(k-1)L] \sim \mathcal{N}(0, \sigma_\theta^2(i-(k-1)L)).
\end{aligned}$$

Henceforth $\mathbb{E} \left[e^{-jU_{\phi_m}} \right] = \varphi_{\phi_m}(-1) = e^{-\frac{\sigma_\phi^2}{2}(i-(k-1)L)}$ and $\mathbb{E} \left[e^{jU_{\theta_k}} \right] = \varphi_{\theta_k}(1) = e^{-\frac{\sigma_\theta^2}{2}(i-(k-1)L)}$, where $\varphi_{\phi_m}(\cdot)$ and $\varphi_{\theta_k}(\cdot)$ are the characteristic functions of U_{ϕ_m} and U_{θ_k} , respectively. The equality (c) follows from (1). \square

In (15), the factor M signifies the combining gain in a coherent receiver (i.e., when $\sigma_\phi = \sigma_\theta = 0$). The factor $e^{-\frac{\sigma_\phi^2 + \sigma_\theta^2}{2}(i-(k-1)L)}$ signifies the loss in effective amplitude gain due to the non-coherency between the received data samples and the estimated channel gains. Note that this non-coherency arises due to the fact that the channel gains for the k -th user are estimated at $t = (k-1)L + l$, $l = 0, \dots, L-1$ and the samples for detecting $x_k[i]$ are received at $t = i + l$, $l = 0, \dots, L-1$, that is, $i - (k-1)L$ samples later. The oscillator phase drift in this time period results in a partial non-coherency. It is clear that the larger this time difference is the smaller the

effective amplitude gain is (the effective amplitude $M\alpha_k e^{-\frac{\sigma_\phi^2 + \sigma_\theta^2}{2}(i-(k-1)L)}$ decreases exponentially with increasing time difference $i - (k-1)L$).

Proposition 2. *The variance $\text{Var}(EN_k[i]) \triangleq \mathbb{E}[|EN_k[i] - \mathbb{E}[EN_k[i]]|^2]$ satisfies, for synchronous operation*

$$\varsigma_k^s[i] \triangleq \text{Var}(EN_k^s[i]) = P_D M^2 \kappa_k[i] + C_k, \quad (16)$$

and for non-synchronous operation

$$\varsigma_k^{ns}[i] \triangleq \text{Var}(EN_k^{ns}[i]) = P_D M^2 \alpha_k^2 \varpi_k[i] + P_D M \xi_k[i] + C_k, \quad (17)$$

where

$$\kappa_k[i] \triangleq \sum_{l=0}^{L-1} \sum_{l'=0}^{L-1} d_{k,l} d_{k,l'} e^{-\sigma_\phi^2 |l-l'|} - \alpha_k^2 e^{-(\sigma_\phi^2 + \sigma_\theta^2)(i-(k-1)L)},$$

$$\xi_k[i] \triangleq \sum_{l=0}^{L-1} \sum_{l'=0}^{L-1} d_{k,l} d_{k,l'} e^{-\sigma_\phi^2 |l-l'|} - \alpha_k^2 e^{-\sigma_\phi^2 (i-(k-1)L)},$$

$$\varpi_k[i] \triangleq e^{-\sigma_\phi^2 (i-(k-1)L)} \left(1 - e^{-\sigma_\theta^2 (i-(k-1)L)} \right),$$

$$C_k \triangleq P_D M \alpha_k \sum_{q=1}^K \alpha_q + \sigma^2 M \left(\frac{P_D}{P_p K} \sum_{q=1}^K \alpha_q + \alpha_k + \frac{\sigma^2}{K P_p} \right).$$

Proof. See the Appendix. □

The second term of the constant C_k in Proposition 2 is the contribution of the additive noise term $AN_k[i]$. This contribution has variance $\mathbb{E}[|AN_k[i]|^2] = \sigma^2 M \left(\frac{P_D}{P_p K} \sum_{q=1}^K \alpha_q + \alpha_k + \frac{\sigma^2}{K P_p} \right)$. The term $\sigma^2 M \frac{P_D}{P_p K} \sum_{q=1}^K \alpha_q$ corresponds to the cross-correlation between the channel estimation error in (5) and the received symbols in (6). The term $\sigma^2 M \alpha_k$ corresponds to the filtered noise (7). Finally, the last term $\sigma^2 M \frac{\sigma^2}{K P_p}$ corresponds to the variance of the channel estimation error.

In the following we provide a coding strategy that justifies the achievable rates we are interested in deriving. From Propositions 1 and 2, it is obvious that $\mathbb{E}[A_k[i]]$ and $\text{Var}(\text{EN}_k[i])$ depend on i and are different for different $i \in \mathcal{I}_d$. Further, for a given i , across multiple transmission blocks, the terms $\mathbb{E}[A_k[i]]$ and $\text{Var}(\text{EN}_k[i])$ are the same and the realizations of $\text{EN}_k[i]$ are i.i.d. Hence, for each i , we have an additive noise SISO channel. This motivates us to consider N_D channel codes for each user, one for each $i \in \mathcal{I}_d$. At the k -th transmitter (user), the symbols of the i -th channel code ($x_k[i]$) are transmitted only during the i -th channel use of each transmission block. Similarly, at the BS, for the k -th user, the i -th received and processed symbols (i.e., $\hat{x}_k[i]$) across different transmission blocks are jointly decoded. Essentially, this implies that, at the BS we have N_D parallel channel decoders for each user. We propose the above scheme of N_D parallel channel codes for each user only to derive a lower bound on the achievable information rate. In practice, due to reasons of complexity, channel coding/decoding would not only be performed across different transmission blocks, but also across consecutive channel uses within each transmission block.

Given the previously described coding strategy, we are now interested in computing a lower bound on the reliable rate of communication for each of the N_D channel codes. Since the data symbols $x_k[i]$ are Gaussian, for each $i \in \mathcal{I}_d$ a lower bound on the information rate for the effective channel in (13) can be computed by considering the worst case (in terms of mutual information) uncorrelated additive noise. With Gaussian information symbols, it is known that the worst case uncorrelated noise is Gaussian with the same variance as that of $\text{EN}_k[i]$ [19]. Consequently, a lower bound on $I(\hat{x}_k[i]; x_k[i])$ (i.e., the mutual information rate for the i -th channel code for user k) is given by Proposition 3.

Proposition 3. *The achievable rate for the i -th channel code for the k -th user is given by*

$$I(\hat{x}_k[i]; x_k[i]) \geq R_k^\times[i] \triangleq \log_2 \left(1 + \frac{P_D M^2 \alpha_k^2 e^{-(\sigma_\phi^2 + \sigma_\theta^2)(i-(k-1)L)}}{\varsigma_k^\times[i]} \right), \quad (18)$$

where $\times = s$ for synchronous operation and $\times = ns$ for non-synchronous operation and ς_k^\times are given in Proposition 2.

Corollary 1. *Based on the lower bounds (18), the proposed TR-MRC receiver exhibits better performance in the case of non-synchronous operation.*

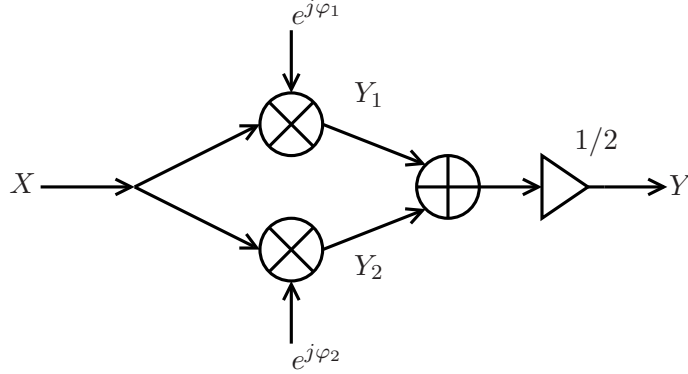


Figure 2: System model for the example.

Proof of Corollary 1.

$$\begin{aligned}
 \varsigma_k^s[i] - \varsigma_k^{n,s}[i] &= P_D M(M-1) \xi_k[i] \\
 &\stackrel{(a)}{\geq} P_D M(M-1) \left(\alpha_k^2 e^{-\sigma_\phi^2(L-1)} - \alpha_k^2 e^{-\sigma_\phi^2(i-(k-1)L)} \right) \\
 &\stackrel{(b)}{\geq} P_D M(M-1) \left(\alpha_k^2 e^{-\sigma_\phi^2(L-1)} - \alpha_k^2 e^{-\sigma_\phi^2(2L-1)} \right) \geq 0.
 \end{aligned}$$

The inequality (a) follows from the fact that $|l - l'| \leq L - 1$ and (1). The inequality (b) follows since $i \geq KL + L - 1 \Rightarrow i - (k-1)L \geq (K-k)L + 2L - 1 \geq 2L - 1$ and $k \leq K$. \square

Note that Corollary 1 compares two lower bounds. However, there are good reasons to expect that these lower bounds are actually quite good predictions of the performance that could be achieved in reality. This is so because substantially we make a Gaussianity assumption on the effective noise. This is also very likely the type of approximation that would be used when deriving a soft decoding (LLR) metric for insertion into for example, a turbo decoder. Hence, using this Gaussian approximation would predict quite well the performance achievable with good channel codes and standard decoding metrics assuming Gaussian noise. Also note that comparing lower bounds that are reasonably tight is a standard practice in the communication theory literature.

Corollary 1 conveys an interesting result that the performance is better when the phase noise processes at the different BS antennas are uncorrelated. However, this is not the first time that such a result is reported.

In [21, Section III.A] the authors study the effect of phase noise in single-user beamforming. The performance measure they use is the error vector magnitude (EVM) and they show that EVM is smallest in the desired direction when uncorrelated phase noise sources are used. In [15, Section VI.D] the authors consider the impact of phase noise distortion in a flat fading channel with maximum ratio combining, using a small phase noise approximation. They also observe that by using separate oscillators the distortion scales as $O(t)$, where t is the time elapsed from channel estimation to data detection. On the other hand, when a common oscillator is used the distortion scales as $O(tM)$. (Note that in contrast to our analysis, [15] used a much simpler model that did not include the effects of intersymbol interference, nor of multiuser interference.) From Corollary 1, it can be argued that the use of independent oscillators at the BS can be beneficial when TR-MRC is used. Also, for a desired sum-rate performance one can choose between a high quality single oscillator or many oscillators of lower quality.

Achievable Sum-Rate

Since no data transmission happens during the training phase, the overall effective information rate achievable by the k -th user is given by,

$$R_k^\times \triangleq \frac{1}{N_c} \sum_{i \in \mathcal{I}_d} R_k^\times [i]. \quad (19)$$

The achievable sum-rate is therefore given by

$$R^\times = \sum_{k=1}^K R_k^\times = \frac{1}{N_c} \sum_{k=1}^K \sum_{i \in \mathcal{I}_d} R_k^\times [i]. \quad (20)$$

It is clear that phase noise degrades the sum-rate performance both with synchronous and non-synchronous operation. To see this formally, note that the sum-rate for the no-phase-noise case can be derived from (18), (19) and (20) by setting $\sigma_\phi^2 = \sigma_\theta^2 = 0$ and is given by

$$\mathcal{R} = \frac{N_D}{N_c} \sum_{k=1}^K \log_2 \left(1 + \frac{P_D M^2 \alpha_k^2}{C_k} \right). \quad (21)$$

Since, $\frac{P_D}{\sigma^2} M^2 \alpha_k^2 \geq \frac{P_D}{\sigma^2} M^2 \alpha_k^2 e^{-(\sigma_\phi^2 + \sigma_\theta^2)(i-(k-1)L)}$ and $\varsigma_k^s [i] \geq \varsigma_k^{ns} [i] \geq C_k$ we have that $\mathcal{R} \geq R^\times$.

4.1 Exact Analysis of Synchronous versus Non-Synchronous Operation for a Toy Channel Model

In the following, we provide a simple example to illustrate that the conclusion drawn from Corollary 1 is the result of a fundamental phenomenon and not an artifact of the techniques used to derive the lower bounds on the information rate. We consider a very simple channel with only phase noise and no AWGN, see Fig. 2. Here $X \in \{\pm 1\}$, $\Pr\{X = +1\} = p$, $\Pr\{X = -1\} = 1 - p$ is the input to the channel. The input X is rotated by φ_1 and φ_2 to form Y_1 and Y_2 , respectively. Let the random variables φ_1, φ_2 model the phase noise, with the following probability mass functions (p.m.f.): $\varphi_i \in \{-\frac{\pi}{2}, 0, \frac{\pi}{2}\}$, $\Pr\{\varphi_i = -\frac{\pi}{2}\} = \Pr\{\varphi_i = 0\} = \Pr\{\varphi_i = \frac{\pi}{2}\} = \frac{1}{3}$, $i = 1, 2$. The output of this discrete memoryless channel (DMC) is given by

$$Y = \frac{1}{2} (e^{j\varphi_1} + e^{j\varphi_2}) X. \quad (22)$$

We now consider two cases, firstly when the two phase noise processes are synchronous (i.e., $\varphi_1 \equiv \varphi_2$) and secondly when they are non-synchronous and mutually independent. In the synchronous case, $\varphi_1 \equiv \varphi_2$ so $Y = e^{j\varphi_1} X$. Then Y takes values in $\mathcal{Y}_s = \{+1, +j, -1, -j\}$. The output symbols have the p.m.f.: $\Pr\{Y = +1\} = p/3$, $\Pr\{Y = -1\} = (1 - p)/3$, $\Pr\{Y = \pm j\} = 1/3$. The capacity of this channel can be calculated as follows

$$\begin{aligned} C_s &= \max_p I(X; Y) = \max_p H(Y) - H(Y|X) \\ &= \max_p \frac{1}{3} H_2(p) = 1/3 \text{ bits,} \end{aligned}$$

where $H_2(p)$ is the binary entropy function.

In the non-synchronous case, where φ_1 and φ_2 are independent of each other, the output variable takes values in $\mathcal{Y}_{ns} = \{+1, \frac{1}{2}(1 + j), \frac{1}{2}(1 - j), j, 0, -j, -\frac{1}{2}(1 - j), -\frac{1}{2}(1 + j), -1\}$. The p.m.f. of the output is $\Pr\{Y = +1\} = p/9$, $\Pr\{Y = (1 \pm j)/2\} = 2p/9$, $\Pr\{Y = \pm j\} = 1/9$, $\Pr\{Y = 0\} = 2/9$, $\Pr\{Y = -(1 \pm j)/2\} = 2(1 - p)/9$, and $\Pr\{Y = -1\} = (1 - p)/9$. We find that $H(Y) = \frac{5}{9} H_2(p) + \log_2 9 - 6/9$ and $H(Y|X = \pm 1) = \log_2 9 - 6/9$. Then, the capacity is given by

$$\begin{aligned}
C_{ns} &= \max_p I(X; Y) = \max_p H(Y) - H(Y|X) \\
&= \max_p \frac{5}{9} H_2(p) = 5/9 \text{ bits.}
\end{aligned}$$

Since $C_s < C_{ns}$, it is concluded that the capacity of the channel in Fig. 2 is strictly larger in the non-synchronous case than in the synchronous case.

Note that the example does not show that the capacity *always* increases if we use independent phase noise sources. However, it shows that *there are* cases where the use of independent phase noise sources can be beneficial.

5 Asymptotic Results

The achievable rates presented in Proposition 3 hold for any M . In this section we present some asymptotic (in M) results based on these achievable rates in order to investigate the Massive MIMO effect in the system under study. In the following $\beta \triangleq \frac{P_p}{P_D} > 0$ denotes the ratio between the per-user average transmit power during the training phase and during the transmission phase.

We first note that in the low SNR regime, the performance loss due to phase noise is not significant. To see this quantitatively, consider the sum-rate when phase noise is present, given by (20). From (18) it is clear that in the low-SNR regime, i.e., when $P_D/\sigma^2 \ll 1$, the dominating factor in the denominator of the argument of the \log_2 function is, in both operation modes, the term $\frac{\sigma^4 M}{K\beta P_D}$. From (21) (after the substitution $P_p = \beta P_D$) it is clear that the term $\frac{\sigma^4 M}{K\beta P_D}$ is also the dominating term in the denominator of the achievable rate expression in the no-phase-noise case. Therefore, the performance loss of both operation modes compared to the no-phase-noise scenario is small. The result is of particular importance since this work focuses mainly on the low SNR (per degree of freedom). This is also often the foreseen operating point of Massive MIMO [5, 22].

We proceed with a result on the sum-rate performance in the high-SNR regime.

Proposition 4. Saturation in the high-SNR regime. *In the presence of phase noise the effective information rate of the k -th user saturates for $\frac{P_D}{\sigma^2} \rightarrow \infty$ to the values, for synchronous operation*

$$R_k^s \rightarrow \frac{1}{N_c} \sum_{i \in \mathcal{I}_d} \log_2 \left(1 + \frac{M\alpha_k^2 e^{-(\sigma_\phi^2 + \sigma_\theta^2)(i-(k-1)L)}}}{M\kappa_k[i] + \alpha_k \sum_{q=1}^K \alpha_q} \right), \quad (23)$$

and for non-synchronous operation

$$R_k^{ns} \rightarrow \frac{1}{N_c} \sum_{i \in \mathcal{I}_d} \log_2 \left(1 + \frac{M\alpha_k^2 e^{-(\sigma_\phi^2 + \sigma_\theta^2)(i-(k-1)L)}}}{M\alpha_k^2 \varpi_k[i] + \xi_k[i] + \alpha_k \sum_{q=1}^K \alpha_q} \right). \quad (24)$$

Proof. The result follows from (18) and the definitions of R_k^s and R_k^{ns} in (19). \square

In the high-SNR regime, MRC is known to be suboptimal since intersymbol interference and multi-user interference dominate the effective noise term. Therefore saturation in the high-SNR regime is observed also in the no-phase-noise case due to the MRC reception strategy.

A particularly desirable property of massive MIMO systems is the array power gain that they offer. The following proposition shows that the phase-noise-impaired single-carrier massive MIMO uplink with TR-MRC receive processing and estimated CSI offers an array gain of $O(\sqrt{M})$ —the same scaling law as for flat fading channels without phase noise, derived in [6].

Proposition 5. *Under the assumptions made in Section 3, an $O(\sqrt{M})$ array gain is achievable.*

Proof. We start by proving the proposition for the synchronous case. Let $P_D = \frac{E_u}{M^\eta}$, where E_u is fixed. Based on the derived achievable rates in Proposition 3, we compute the maximum possible exponent, $\eta > 0$, such that a fixed, non-zero rate for the i -th code of user k can be achieved, while the transmit power of each user is scaled as $1/M^\eta$ with increasing M . From argument of the log expression in (18), i.e. the effective SINR, we have

$$\begin{aligned} \text{SINR}_k[i] &= \frac{\frac{E_u M \alpha_k^2}{\sigma^2 M^\eta} e^{-(\sigma_\phi^2 + \sigma_\theta^2)(i-(k-1)L)}}{\frac{E_u M \kappa_k[i]}{\sigma^2 M^\eta} + \frac{E_u \alpha_k \sum_q \alpha_q}{\sigma^2 M^\eta} + \alpha_k + \frac{\sum_q \alpha_q}{\beta K} + \frac{M^\eta \sigma^2}{K \beta E_u}} \\ &= \frac{\frac{E_u \alpha_k^2}{\sigma^2} e^{-(\sigma_\phi^2 + \sigma_\theta^2)(i-(k-1)L)}}{\frac{E_u \kappa_k[i]}{\sigma^2} + \frac{E_u \alpha_k \sum_q \alpha_q}{M \sigma^2} + M^{\eta-1} \left(\alpha_k + \frac{\sum_q \alpha_q}{\beta K} \right) + \frac{M^{2\eta-1} \sigma^2}{K \beta E_u}}. \end{aligned}$$

6. Impact of Phase Noise Separately at the BS and at the User Terminals 97

As $M \rightarrow \infty$ we have $\lim_{M \rightarrow \infty} R_k^s[i] > 0$ if $\eta - 1 \leq 0$ and $2\eta - 1 \leq 0 \Rightarrow \eta \leq 1/2$. For $\eta = 1/2$ the rate R_k^s converges to the value (as $M \rightarrow \infty$)

$$R_k^s \rightarrow \frac{1}{N_c} \sum_{i \in \mathcal{I}_d} \log_2 \left(1 + \frac{\frac{E_u}{\sigma^2} \alpha_k^2 e^{-(\sigma_\phi^2 + \sigma_\theta^2)(i-(k-1)L)}}{\frac{E_u}{\sigma^2} \kappa_k[i] + \frac{\sigma^2}{K\beta E_u}} \right). \quad (25)$$

Similarly, it can be proved that the array gain for the non-synchronous operation is $O(\sqrt{M})$ and the rate approaches (as $M \rightarrow \infty$) the value

$$R_k^{ns} \rightarrow \frac{1}{N_c} \sum_{i \in \mathcal{I}_d} \log_2 \left(1 + \frac{\frac{E_u}{\sigma^2} \alpha_k^2 e^{-(\sigma_\phi^2 + \sigma_\theta^2)(i-(k-1)L)}}{\frac{E_u}{\sigma^2} \alpha_k^2 \varpi_k[i] + \frac{\sigma^2}{K\beta E_u}} \right). \quad (26)$$

It is clear that for $\eta > 1/2$ the achievable rates approach 0 as $M \rightarrow \infty$. \square

6 Impact of Phase Noise Separately at the BS and at the User Terminals

Based on the preceding analysis, we examine two special cases of particular interest. Namely, we study the impact on sum-rate performance, when there is phase noise only at the user terminals (UTs) and not at the BS (i.e. $\sigma_\phi^2 = 0$ and $\sigma_\theta^2 \neq 0$) and vice versa (i.e. $\sigma_\phi^2 \neq 0$ and $\sigma_\theta^2 = 0$).

6.1 Special Case 1: Phase Noise Only at the UTs, $\sigma_\phi^2 = 0$

If the oscillators at the BS are ideal, there is no distinction between synchronous and non-synchronous operation. From (18) it follows immediately that the lower bound in this case is given by

$$R_k[i] = \log_2 \left(1 + \frac{\frac{P_D M \alpha_k^2 e^{-\sigma_\theta^2(i-(k-1)L)}}{\sigma^2}}{\frac{P_D M}{\sigma^2} \alpha_k^2 \left(1 - e^{-\sigma_\theta^2(i-(k-1)L)} \right) + \frac{C_k}{\sigma^2 M}} \right). \quad (27)$$

In the high SNR limit the rate saturates at the value

$$R_k[i] \rightarrow \log_2 \left(1 + \frac{M \alpha_k e^{-\sigma_\theta^2(i-(k-1)L)}}{M \alpha_k \left(1 - e^{-\sigma_\theta^2(i-(k-1)L)} \right) + \sum_{q=1}^K \alpha_q} \right). \quad (28)$$

Further, by scaling the transmit power as $P_D = E_u/\sqrt{M}$ we have the limiting expression as $M \rightarrow \infty$

$$R_k[i] \rightarrow \log_2 \left(1 + \frac{\frac{E_u}{\sigma^2} \alpha_k^2 e^{-\sigma_\theta^2(i-(k-1)L)}}{\frac{E_u}{\sigma^2} \alpha_k^2 \left(1 - e^{-\sigma_\theta^2(i-(k-1)L)} \right) + \frac{\sigma^2}{K\beta E_u}} \right). \quad (29)$$

The expressions in (27), (28) and (29) are qualitatively similar to the case of synchronous operation at the BS. In the following we provide an intuitive explanation of this similarity. Consider the link between user k and the BS. Irrespectively of whether there is phase noise at the BS or not, the distortion in the received signal at each BS antenna due to the phase noise at the user adds up after TR-MRC processing, giving an additional interference term (see $\text{IF}_k[i]$ in (14)) with a standard deviation that scales as $O(M)$.

6.2 Special Case 2: Phase Noise Only at the BS, ($\sigma_\phi^2 \neq 0$ and $\sigma_\theta^2 = 0$)

In this case the achievable rate for the synchronous case is given by

$$R_k^s[i] = \log_2 \left(1 + \frac{\frac{P_D M \alpha_k^2}{\sigma^2} e^{-\sigma_\phi^2(i-(k-1)L)}}{\frac{P_D M}{\sigma^2} \xi_k[i] + \frac{C_k}{\sigma^2 M}} \right), \quad (30)$$

and for the non-synchronous case

$$R_k^{ns}[i] = \log_2 \left(1 + \frac{\frac{P_D M}{\sigma^2} \alpha_k^2 e^{-\sigma_\phi^2(i-(k-1)L)}}{\frac{P_D}{\sigma^2} \xi_k[i] + \frac{C_k}{\sigma^2 M}} \right). \quad (31)$$

In the high SNR regime the above rates saturate at the following values

$$R_k^s[i] \rightarrow \log_2 \left(1 + \frac{M \alpha_k^2 e^{-\sigma_\phi^2(i-(k-1)L)}}{M \xi_k[i] + \alpha_k \sum_{q=1}^K \alpha_q} \right) \quad (32)$$

$$R_k^{ns}[i] \rightarrow \log_2 \left(1 + \frac{M \alpha_k^2 e^{-\sigma_\phi^2(i-(k-1)L)}}{\xi_k[i] + \alpha_k \sum_{q=1}^K \alpha_q} \right). \quad (33)$$

Further, by scaling the transmit power as $P_D = E_u/\sqrt{M}$ we have the limiting expressions as $M \rightarrow \infty$ for the synchronous operation

$$R_k^s[i] \rightarrow \log_2 \left(1 + \frac{\frac{E_u}{\sigma^2} \alpha_k^2 e^{-\sigma_\phi^2(i-(k-1)L)}}{\frac{E_u}{\sigma^2} \xi_k[i] + \frac{\sigma^2}{K\beta E_u}} \right), \quad (34)$$

and for the non-synchronous operation

$$R_k^{ns}[i] \rightarrow \log_2 \left(1 + \left(\frac{E_u}{\sigma^2} \right)^2 K\beta \alpha_k^2 e^{-\sigma_\phi^2(i-(k-1)L)} \right). \quad (35)$$

The expressions in (30), (32) and (34) are qualitatively similar to the case of phase noise only at the user terminals and to the general case with synchronous operation at the BS. In fact, it is the symmetric case as in Section 6.1. This behavior can be explained by arguments similar to the ones used there.

However, in the expressions for the non-synchronous operation (31), (33) and (35) we observe a fundamentally different behavior. Firstly, in (33) we note that by increasing the number of BS antennas, we can increase the high-SNR saturation value of the achievable rate arbitrarily. In addition, from (35) it is clear that in the large array regime we can arbitrarily increase the limiting expression by appropriately selecting the value E_u . These observations lead to the conclusion that the distortions introduced by independent oscillators at the BS asymptotically vanish, when TR-MRC reception is used. We remark that similar behavior was also noted in [15], where the authors demonstrate that the dominating impairment is the one at the hardware of the user equipment, while impairments at the BS from independent sources asymptotically vanish as $M \rightarrow \infty$.

7 Numerical Examples

In this section, we present numerical examples of the main results presented in Sections 4–6. Throughout the section we selected $T_s = 0.1\mu s$ and $f_c = 2$ GHz, which correspond to typical values of wideband wireless communication systems, such as the WLAN IEEE 802.11. The reference value of the oscillator parameter c_ϕ (and c_θ) is set to $c_\phi = 4.7 \times 10^{-18}(\text{rad Hz})^{-1}$, which also corresponds to a typical oscillator in WLAN IEEE 802.11 equipment [23, Table 1]. However, we will refer to the standard deviation of the

phase noise innovations, i.e. σ_ϕ and σ_θ , since this is a more intuitive measure of the oscillator quality. For the parameters selected above and the relations in Section 2.1, $\sigma_\phi = 0.49^\circ$. In typical cellular systems the delay spread is of the order of microseconds. We select $L = 20$, which corresponds to $2\mu s$ of delay spread for the selected symbol rate. We selected the large scale fading factors as $\alpha_k = 1, \forall k \in \{1, \dots, K\}$, since the main purpose of this work is to understand the effect of phase noise and not of large scale fading. However, the same relations can be used with other choices of α_k 's, when the study of particular propagation conditions is of interest. Further, we have selected a common power delay profile of every user as $d_{k,l} = e^{-0.35l} / \sum_{p=0}^{L-1} e^{-0.35p}$, $l = \{0, \dots, L-1\}$. We note that the power delay profile enters the rate expressions through the terms $\kappa_k[i]$ and $\xi_k[i]$ (see Proposition 2). For most reasonable choices of σ_ϕ the choice of a particular PDP has a negligible effect on the achievable sum-rate. This choice of PDP and large scale fading is the same for all the figures that follow.

In Fig. 3 the sum-rate performance of the system, as given by (20), is plotted as a function of $\frac{P_D}{\sigma^2}$ for $N_D = 1000$ with $M = 200, K = 10$. The sum-rate achieved without phase noise (21) is plotted for the sake of comparison. We observe that at low SNR, the loss in sum-rate performance is insignificant. This observation supports our argument on the low SNR performance at the beginning of Section 5. We plot the sum-rate as a function of $\frac{P_D}{\sigma^2}$ for various choices of σ_ϕ and σ_θ . It is clear from Fig. 3b that when the phase noise at the user terminals is dominant both operation modes have similar performance. On the other hand, when the phase noise at the BS is dominant, as in Fig. 3c, the sum-rate of the non-synchronous operation is significantly higher than the synchronous operation mode. This is in agreement with the discussion in Section 6.

A significant desirable property of massive MIMO systems is the array power gain that they offer, facilitating the design of highly power-efficient communication systems [3, 6, 24]. Proposition 5 extends this result to the case of single-carrier frequency-selective Massive MU-MIMO systems impaired with phase noise. The above observation is further supported through Fig. 4, where the minimum per-user $\frac{P_D}{\sigma^2}$ required to achieve a fixed per-user information rate of $r = 2$ bpcu is plotted as a function of the number of BS antennas for $N_D = 1000$ and $K = 10$ for $\sigma_\phi = \sigma_\theta = 0.49^\circ$. The plot for the phase-noise-free case is also given for the sake of comparison. We observe that by doubling the number of BS antennas we can reduce the per-user required $\frac{P_D}{\sigma^2}$ by 1.5dB, for sufficiently large M . This illustrates the validity of Proposition 5.

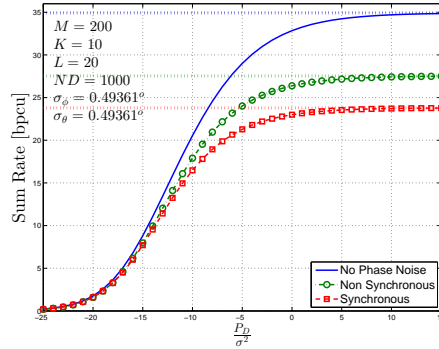
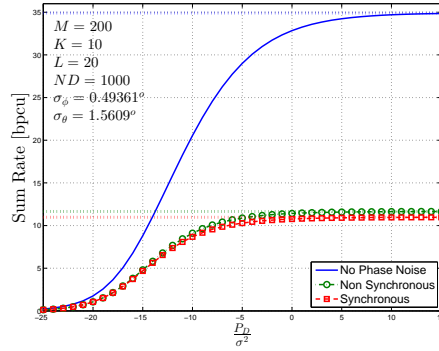
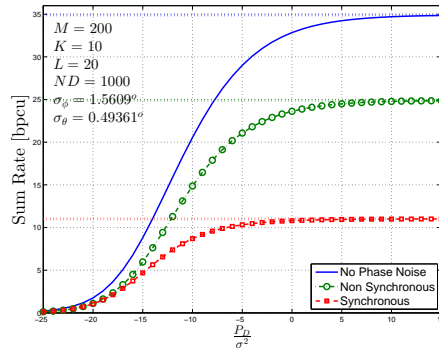
(a) $\sigma_\phi = \sigma_\theta = 0.49^\circ$ (b) $\sigma_\phi = 0.49^\circ, \sigma_\theta = 1.56^\circ$ (c) $\sigma_\phi = 1.56^\circ, \sigma_\theta = 0.49^\circ$

Figure 3: Sum-rate as a function of $\frac{P_D}{\sigma^2}$ [dB] for $M = 200$, $K = 10$, $L = 20$ and $N_D = 1000$. The dotted vertical lines denote the high SNR asymptotic values of the achievable sum-rates.

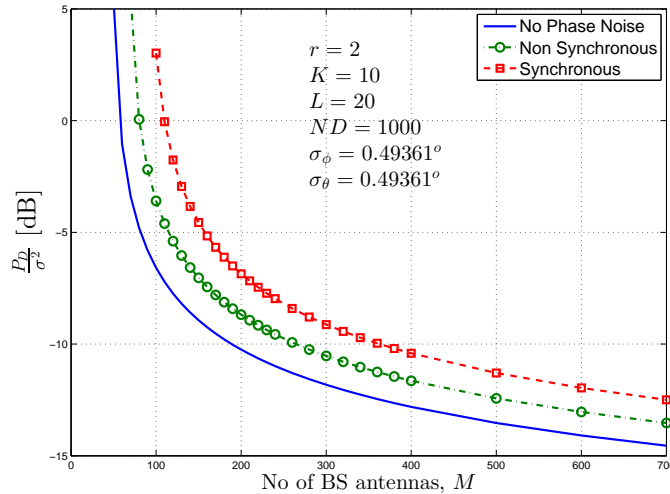


Figure 4: Minimum required $\frac{P_D}{\sigma_D^2}$ to achieve a fixed per-user information rate of $r = 2$ bpcu as a function of M for fixed $K = 10$ users, $\sigma_\phi = \sigma_\theta = 0.49^\circ$ and $N_D = 1000$.

From Fig. 4 we are motivated to study the gap in required $\frac{P_D}{\sigma_D^2}$ between the phase-noise-impaired cases and the no-phase-noise operation. In Table 1 we present numerical results on this gap. Each row corresponds to a different oscillator constant $c_\phi = c_\theta$, namely, 9.4×10^{-19} , 4.7×10^{-18} and $2.35 \times 10^{-17}(\text{rad Hz})^{-1}$, which correspond to standard deviation of phase noise innovations of 0.22° , 0.49° and 1.1° , respectively. In order to give a more intuitive measure of the disturbance introduced by phase noise, we list the vertical $\frac{P_D}{\sigma_D^2}$ gap as a function of the standard deviation of the accumulated phase noise drift at a time difference of $N_D + L - 1$ channel uses (i.e., the time difference between the end of the training phase and the end of the data phase). This result is shown in Table 1. As expected, the performance gap is minimal for small phase noise drift and increases as the standard deviation of the phase noise drift increases.

It is also interesting to study the gap in required $\frac{P_D}{\sigma_D^2}$ as a function of the desired per-user information rate. For this purpose we provide Table 2. There, we tabulate the gap in required $\frac{P_D}{\sigma_D^2}$ in dB for various values of the per-user desired information rate for the synchronous and non-synchronous mode, for $N_D = 1000$ channel uses, $\sigma_\phi = \sigma_\theta = 0.49^\circ$, $K = 10$ users and $M = 500$ BS antennas. In the low spectral efficiency regime this gap is

Table 1: Gap in required $\frac{P_D}{\sigma^2}$ due to phase noise for $N_D = 1000$ and a fixed per-user information rate $r = 1$ bpcu. The number of users is fixed to $K = 10$.

$\sigma_\phi \sqrt{N_D}$ (degrees)	Gap in required $\frac{P_D}{\sigma^2}$ [dB]			
	Synchronous		Non-Synchronous	
	M=500	M=2500	M=500	M=2500
7.05°	0.1174	0.1055	0.0828	0.0744
15.76°	0.6145	0.5492	0.4192	0.3753
35.23°	4.7459	3.9629	2.3071	2.0116

minimal. However, as the desired per-user information rate increases the gap increases at a faster rate. When the desired per-user information rate increases from 2 bpcu to 2.5 bpcu, which corresponds to 25% increase, the gap in dB in the case of non-synchronous operation doubles, whereas in the synchronous operation mode the vertical gap increases more than two times. This happens because the desired per-user rate is close to the high-SNR saturation rate for the case of synchronous receivers⁵. As a result, a large increase in the transmit power is required in order to achieve the desired information rate.

Table 2: Gap in required $\frac{P_D}{\sigma^2}$ due to phase noise for $N_D = 1000$, $\sigma_\phi = \sigma_\theta = 0.49^\circ$, $K = 10$ users and $M = 500$ BS antennas for various values of the desired per-user information rate in bits per channel use [bpcu].

Per-user rate	Gap in required $\frac{P_D}{\sigma^2}$ [dB]	
	Synchronous	Non-Synchronous
0.25	0.2768	0.2481
0.5	0.3625	0.2941
1	0.6145	0.4192
2	2.2356	1.0987
2.5	6.8694	2.1749

⁵With the selected parameters, the high-SNR saturation value for the synchronous operation is 2.66 bpcu per user.

For fixed M , K and L there is a fundamental trade-off between the length of the data interval, N_D , and the achievable sum-rate performance. A fraction $\frac{KL}{N_c}$ of each coherence interval is spent on training. Since a fixed time interval of KL channel uses is required for channel estimation, a small data interval, N_D , leads to underutilization of the available resources, yielding a low sum-rate performance. As N_D increases, more resources are utilized for the data transmission, increasing the sum-rate performance. However, as it can be seen from (18), $R_k^s[i] < R_k^s[i-1]$ and $R_k^{ns}[i] < R_k^{ns}[i-1]$, which implies that the gain of increasing the data interval diminishes with increasing N_D . In fact, the individual rates $R_k^s[i]$ and $R_k^{ns}[i]$ approach 0 as $i \rightarrow \infty$. This phenomenon occurs because with large N_D , the phase noise drift in the oscillators is so large such that there is a total loss of coherency between the received symbols during the data phase and the estimated channel at the beginning of the transmission block.

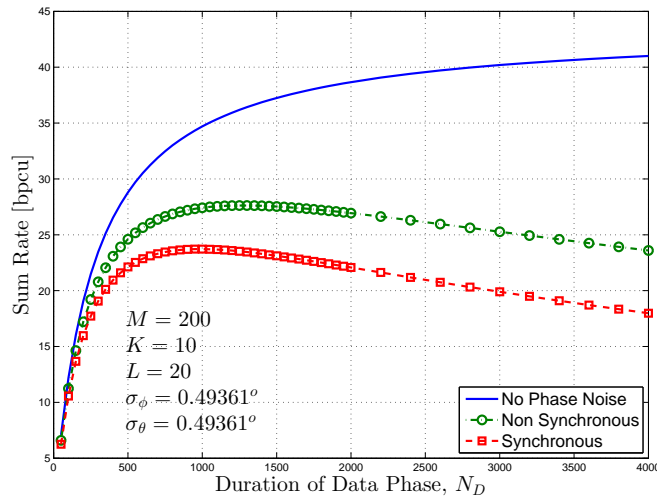


Figure 5: Sum-rate performance as a function of N_D , with fixed, $\sigma_\phi = \sigma_\theta = 0.49^\circ$, $\frac{P_D}{\sigma^2} = 10$ dB, $M = 200$ BS antennas, $K = 10$ users and $L = 20$ taps.

In Fig. 5 the sum-rate performance is plotted as a function of N_D for $\sigma_\phi = \sigma_\theta = 0.49^\circ$. In the no-phase-noise case the optimal value of N_D is infinity. However, there is a clear trade-off between the sum-rate and the length of the data interval in the phase-noise-impaired operation modes.

Further insight can be obtained by considering the optimum number of scheduled users. In practice, the coherence interval is finite and therefore

the training overhead upper-bounds the optimum number of scheduled users. Now, consider the case where the coherence interval is arbitrarily long. Then for the no-phase noise case, the optimal N_D is unbounded. In that case one can increase the number of users, thereby achieving an increase in the sum-rate performance due to the spatial multiplexing of more users in the same time-frequency resource. In the presence of phase noise increasing the number of scheduled users, K , not only increases the length of the training overhead, but it also increases the phase drift between the estimated channel coefficients and the actual realizations of the effective channel impulse responses during the data interval. That is, by increasing the number of users, K , the partial loss of coherency between the estimated channel coefficients and the actual effective channels during data transmission is also increased. As a result, with increasing K the increase in the achievable sum-rate during the data interval may eventually become insignificant to compensate for the reduction in sum-rate due to this partial loss of coherency. In Fig. 6, for every K the maximum achievable sum-rate performance is found by maximizing with respect to N_D and, subsequently, this maximum sum-rate performance is plotted as a function of K for $\frac{P_D}{\sigma^2} = 10$ dB, $M = 200$ BS antennas and $L = 20$ taps for the no phase noise case, the synchronous operation mode and the non-synchronous operation mode. It is clear that the sum-rate performance is not monotonically increasing in the phase-noise-impaired cases as it is in the no phase noise case. However, it has a unimodal shape. This implies that in practice the optimum number of scheduled users is not only upper-bounded by the length of the coherence interval, but it is also upper-bounded as a consequence of the phase noise.

8 Conclusions

Phase noise is an inevitable hardware impairment in communication systems. We studied the effect of phase noise on the sum-rate performance of single-carrier transmission in a MU-MIMO uplink with an excess of BS antennas. Two distinct operation modes in terms of the phase noise processes at the BS antennas are considered, namely, synchronous and non-synchronous operation. Since the knowledge of the exact channel realizations is not available, CSI is acquired via uplink training. The BS uses TR-MRC receive processing to detect the information symbols. An analytical

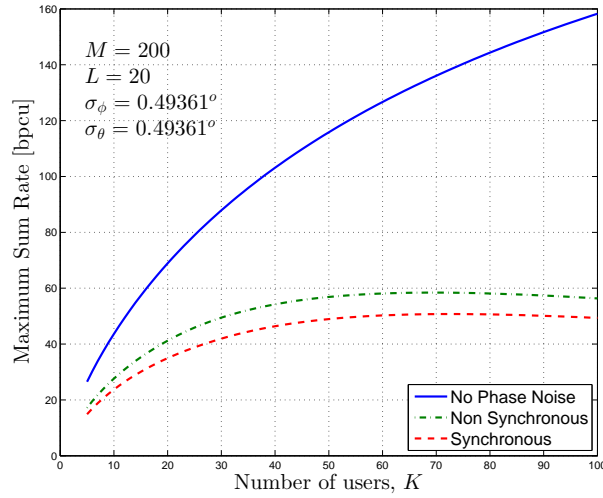


Figure 6: Maximum sum-rate performance as a function of K , with fixed $\frac{P_D}{\sigma^2} = 10$ dB, $\sigma_\phi = \sigma_\theta = 0.49^\circ$, $M = 200$ BS antennas and $L = 20$ taps. For each K , N_D is optimally chosen.

expression for the achievable sum-rate is rigorously derived for both operation modes. Based on the derived achievable sum-rates, we observe that it can be beneficial to use independent instead of fully synchronous phase noise sources. It is also shown that at low SNR, phase noise has little impact on the sum-rate performance. Further, the proposed receive processing achieves an $O(\sqrt{M})$ array power gain, extending earlier results where phase noise was not considered. Finally, due to the progressive phase noise drift in the oscillators, there is a fundamental trade-off between the length of the time interval used for data transmission and the sum-rate performance.

Appendix

In this appendix we state the proof of Proposition 2. For both operation modes, we have

$$\begin{aligned} \text{Var}(\text{EN}_k[i]) &\triangleq \mathbb{E} [|\text{EN}_k[i] - \mathbb{E}[\text{EN}_k[i]]|^2] = \text{Var}(\text{IF}_k[i]) \\ &+ \text{Var}(\text{ISI}_k[i]) + \text{Var}(\text{MUI}_k[i]) + \text{Var}(\text{AN}_k[i]) \end{aligned}$$

since the terms in $\text{EN}_k[i]$ are mutually uncorrelated. We start by computing the terms $\text{Var}(\text{ISI}_k[i])$, $\text{Var}(\text{MUI}_k[i])$, $\text{Var}(\text{AN}_k[i])$ for the non-synchronous case, which are the same for both operation modes and conclude with the term $\text{Var}(\text{IF}_k[i])$, the calculation of which is different depending on the operation mode. First we compute the variance of the ISI term.

$$\begin{aligned} \mathbb{E}[|\text{ISI}_k[i]|^2] &= \mathbb{E}[\sqrt{P_D} \sum_{m=1}^M \sum_{l=0}^{L-1} \sum_{\substack{q=0 \\ q \neq l}}^{L-1} g_{m,k,l}^* g_{m,k,q} \vartheta_{i,l,p}^{(m,k,k)} \\ &\cdot x_k[i+l-q]|^2] = P_D \sum_{m=1}^M \sum_{m'=1}^M \sum_{l=0}^{L-1} \sum_{\substack{l'=0 \\ l' \neq l}}^{L-1} \sum_{\substack{p=0 \\ p \neq l}}^{L-1} \sum_{\substack{p'=0 \\ p' \neq l'}}^{L-1} \\ &\cdot \mathbb{E}[g_{m,k,l}^* g_{m,k,p} g_{m',k,p'}^* g_{m',k,l'}] \\ &\cdot \mathbb{E}\left[e^{-j(\phi_m[i+l] - \phi_{m'}[i+l'] - \phi_m[(k-1)L+l] + \phi_{m'}[(k-1)L+l'])}\right] \\ &\cdot \mathbb{E}\left[e^{j(\theta_k[i+l-p] - \theta_k[(k-1)L] - \theta_k[i+l'-p'] + \theta_k[(k-1)L])}\right] \\ &\cdot \mathbb{E}[x_k[i+l-p] x_k^*[i+l'-p']] = P_D \sum_{m=1}^M \sum_{l=0}^{L-1} \sum_{\substack{q=0 \\ q \neq l}}^{L-1} d_{k,l} d_{k,q} \\ &= P_D M \left(\alpha_k^2 - \sum_{l=0}^{L-1} d_{k,l}^2 \right), \end{aligned}$$

where we have used the fact that the channel coefficients, the phase noise processes and the data symbols are mutually independent. The last step follows from the normalization of the PDP (see (1)). We will make use of these facts in all the following derivations as well. We proceed with the calculation of the multi-user interference.

$$\begin{aligned}
 \mathbb{E}[|\text{MUI}_k[i]|^2] &= \mathbb{E}\left[\left|\sqrt{P_D} \sum_{\substack{m=1 \\ q \neq k}}^M \sum_{q=1}^K \sum_{l=0}^{L-1} \sum_{p=0}^{L-1} g_{m,k,l}^* g_{m,q,p} \vartheta_{i,l,p}^{(m,k,q)}\right.\right. \\
 &\quad \left.\left.\cdot x_q[i+l-p]\right|^2\right] = P_D \sum_{m=1}^M \sum_{\substack{m'=1 \\ q \neq k}}^M \sum_{q=1}^K \sum_{\substack{q'=1 \\ q' \neq k}}^K \sum_{l=0}^{L-1} \sum_{\substack{l'=0 \\ p \neq l}}^{L-1} \sum_{\substack{p=0 \\ p' \neq l}}^{L-1} \sum_{\substack{p'=0 \\ p' \neq l}}^{L-1} \\
 &\quad \cdot \mathbb{E}\left[g_{m,k,l}^* g_{m,q,p} g_{m',q',p'}^* g_{m',k,l'}\right] \\
 &\quad \cdot \mathbb{E}\left[e^{-j(\phi_m[i+l] - \phi_{m'}[i+l'] - \phi_m[(k-1)L+l] + \phi_{m'}[(k-1)L+l'])}\right] \\
 &\quad \cdot \mathbb{E}\left[e^{j(\theta_q[i+l-p] - \theta_k[(k-1)L] - \theta_{q'}[i+l'-p'] + \theta_k[(k-1)L])}\right] \\
 &\quad \cdot \mathbb{E}\left[x_q[i+l-p] x_{q'}^*[i+l'-p']\right] \\
 &= P_D \sum_{\substack{m=1 \\ q \neq k}}^M \sum_{q=1}^K \sum_{l=0}^{L-1} \sum_{p=0}^{L-1} d_{k,l} d_{q,p} = P_D M \alpha_k \sum_{\substack{q=1 \\ q \neq k}}^K \alpha_q
 \end{aligned}$$

We conclude the first part of the proof with the calculation of the variance of the additive noise term.

$$\begin{aligned}
 \mathbb{E}[|\text{AN}_k[i]|^2] &= \mathbb{E}\left[\left|\sqrt{\frac{P_D}{P_p K L}} \sum_{m=1}^M \sum_{q=1}^K \sum_{l=0}^{L-1} \sum_{p=0}^{L-1} g_{m,q,p}\right.\right. \\
 &\quad \left.\left.\cdot e^{-j(\phi_m[i+l] - \theta_q[i+l-p])} n_m[(k-1)L+l] x_q[i+l-p]\right|^2\right] \\
 &\quad + \mathbb{E}\left[\left|\sum_{m=1}^M \sum_{l=0}^{L-1} \hat{g}_{m,k,l}^* n_m[i+l]\right|^2\right] \\
 &= \frac{P_D}{P_p K L} \sum_{m=1}^M \sum_{\substack{m'=1 \\ q \neq k}}^M \sum_{q=1}^K \sum_{\substack{q'=1 \\ q' \neq k}}^K \sum_{l=0}^{L-1} \sum_{\substack{l'=0 \\ p \neq l}}^{L-1} \sum_{\substack{p=0 \\ p' \neq l}}^{L-1} \sum_{\substack{p'=0 \\ p' \neq l}}^{L-1}
 \end{aligned}$$

$$\begin{aligned}
& \mathbb{E}[(g_{m,q,p} e^{-j(\phi_m[i+l]-\theta_q[i+l-p])} n_m[(k-1)L+l] x_q[i+l-p]) \\
& \cdot (g_{m',q',p'} e^{-j(\phi_{m'}[i+l']-\theta_{q'}[i+l'-p'])} n_{m'}[(k-1)L+l']) \\
& \cdot x_{q'}[i+l'-p'])^*] + \sigma^2 \sum_{m=1}^M \sum_{l=0}^{L-1} \mathbb{E}[|\hat{g}_{m,k,l}|^2] \\
& = \frac{P_D \sigma^2}{P_p K L} \sum_{m=1}^M \sum_{q=1}^K \sum_{l=0}^{L-1} \sum_{\substack{a=1-L \\ 0 \leq l-a \leq L-1}}^{L-1} d_{q,l-a} \\
& + \sigma^2 \sum_{m=1}^M \sum_{l=0}^{L-1} \left(\frac{\sigma^2}{P_p K L} + \mathbb{E}[|g_{m,k,l}|^2] \right) \\
& = \sigma^2 M \left(\frac{P_D}{P_p K} \sum_{q=1}^K \alpha_q + \frac{\sigma^2}{P_p K} + \alpha_k \right)
\end{aligned}$$

We proceed by calculating the variance of the term $\text{IF}_k[i]$. It holds

$$\begin{aligned}
\text{Var}(\text{IF}_k[i]) &= \mathbb{E}[|(A_k[i] - \mathbb{E}[A_k[i]])x_k[i]|^2] \\
&= \mathbb{E}[|A_k[i]|^2] - |\mathbb{E}[A_k[i]|]^2.
\end{aligned}$$

Based on the result of Proposition 1 it is sufficient to calculate $\mathbb{E}[|A_k[i]|^2]$ for each operation mode. We start with the synchronous operation.

$$\begin{aligned}
 \mathbb{E} \left[|A_k[i]|^2 \right] &= P_D \sum_{m=1}^M \sum_{l=0}^{L-1} \mathbb{E}[|g_{m,k,l}|^4] \\
 &+ P_D \sum_{m=1}^M \sum_{l=0}^{L-1} \sum_{\substack{l'=0 \\ l' \neq l}}^{L-1} \mathbb{E}[|g_{m,k,l}|^2] \mathbb{E}[|g_{m,k,l'}|^2] \\
 &\cdot \mathbb{E}[e^{-j(\phi[i+l]-\phi[i+l']-\phi[(k-1)L+l]+\phi[(k-1)L+l'])}] \\
 &+ P_D \sum_{m=1}^M \sum_{\substack{m'=1 \\ m' \neq m}}^M \sum_{l=0}^{L-1} \sum_{l'=0}^{L-1} \mathbb{E}[|g_{m,k,l}|^2] \mathbb{E}[|g_{m',k,l'}|^2] \\
 &\cdot \mathbb{E}[e^{-j(\phi[i+l]-\phi[i+l']-\phi[(k-1)L+l]+\phi[(k-1)L+l'])}] \\
 &= P_D M \sum_{l=0}^{L-1} 2d_{k,l}^2 + P_D M \sum_{l=0}^{L-1} \sum_{\substack{l'=0 \\ l' \neq l}}^{L-1} d_{k,l} d_{k,l'} e^{-\sigma_\phi^2 |l-l'|} \\
 &+ P_D M (M-1) \sum_{l=0}^{L-1} \sum_{l'=0}^{L-1} d_{k,l} d_{k,l'} e^{-\sigma_\phi^2 |l-l'|} \\
 &= P_D M \sum_{l=0}^{L-1} d_{k,l}^2 + P_D M^2 \sum_{l=0}^{L-1} \sum_{l'=0}^{L-1} d_{k,l} d_{k,l'} e^{-\sigma_\phi^2 |l-l'|}
 \end{aligned}$$

Finally, for the synchronous operation, the effective noise variance, is given by

$$\varsigma_k^s[i] \triangleq \text{Var}(\text{EN}_k^s[i]) = P_D M^2 \kappa_k[i] + C_k.$$

We conclude with the calculation of the term $\mathbb{E} \left[|A_k[i]|^2 \right]$ for the non-

synchronous mode.

$$\begin{aligned}
\mathbb{E} \left[|A_k[i]|^2 \right] &= P_D \sum_{m=1}^M \sum_{l=0}^{L-1} \mathbb{E}[|g_{m,k,l}|^4] \\
&+ P_D \sum_{m=1}^M \sum_{l=0}^{L-1} \sum_{\substack{l'=0 \\ l' \neq l}}^{L-1} \mathbb{E}[|g_{m,k,l}|^2] \mathbb{E}[|g_{m,k,l'}|^2] \\
&\cdot \mathbb{E}[e^{-j(\phi_m[i+l] - \phi_m[i+l'] - \phi_m[(k-1)L+l] + \phi_m[(k-1)L+l'])}] \\
&+ P_D \sum_{m=1}^M \sum_{\substack{m'=1 \\ m' \neq m}}^M \sum_{l=0}^{L-1} \sum_{l'=0}^{L-1} \mathbb{E}[|g_{m,k,l}|^2] \mathbb{E}[|g_{m',k,l'}|^2] \\
&\cdot \mathbb{E}[e^{-j(\phi_m[i+l] - \phi_{m'}[i+l'] - \phi_m[(k-1)L+l] + \phi_{m'}[(k-1)L+l'])}] \\
&= P_D M \sum_{l=0}^{L-1} 2d_{k,l}^2 + P_D M \sum_{l=0}^{L-1} \sum_{\substack{l'=0 \\ l' \neq l}}^{L-1} d_{k,l} d_{k,l'} e^{-\sigma_\phi^2 |l-l'|} \\
&+ P_D M (M-1) \sum_{l=0}^{L-1} \sum_{l'=0}^{L-1} d_{k,l} d_{k,l'} e^{-\sigma_\phi^2 (i-(k-1)L)} \\
&= P_D M \sum_{l=0}^{L-1} d_{k,l}^2 + P_D M \sum_{l=0}^{L-1} \sum_{l'=0}^{L-1} d_{k,l} d_{k,l'} e^{-\sigma_\phi^2 |l-l'|} \\
&+ P_D M (M-1) \alpha_k^2 e^{-\sigma_\phi^2 (i-(k-1)L)}.
\end{aligned}$$

The variance for the non-synchronous operation is

$$\varsigma_k^{ns}[i] \triangleq \text{Var}(\text{EN}_k^{ns}[i]) = P_D M \xi_k[i] + P_D M^2 \varpi_k[i] + C_k.$$

Bibliography

- [1] G. Foschini and M. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, pp. 311–335, Mar. 1998.
- [2] D. Gesbert, M. Kountouris, R. W. Heath Jr., C.-B. Chae, and T. Sälzer, "Shifting the MIMO Paradigm," *IEEE Signal Processing Magazine*, vol. 24, pp. 36–46, Sept. 2007.
- [3] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, pp. 3590–3600, Nov. 2010.
- [4] F. Rusek, D. Persson, B. K. Lau, E. Larsson, T. Marzetta, O. Edfors, and F. Tufvesson, "Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays," *IEEE Signal Processing Magazine*, vol. 30, pp. 40–60, Jan. 2013.
- [5] E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta, "Massive MIMO for Next Generation Wireless Systems," *IEEE Communications Magazine*, vol. 52, pp. 186–195, Feb. 2014.
- [6] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Transactions on Communications*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.
- [7] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of Cellular Networks: How Many Antennas Do We Need?," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 160–171, Feb. 2013.
- [8] T. Pollet, M. Van Bladel, and M. Moeneclaey, "BER sensitivity of OFDM systems to carrier frequency offset and Wiener phase

- noise," *IEEE Transactions on Communications*, vol. 43, pp. 191–193, Feb/Mar/Apr 1995.
- [9] L. Tomba, "On the effect of Wiener phase noise in OFDM systems," *IEEE Transactions on Communications*, vol. 46, pp. 580–583, May 1998.
- [10] S. Wu and Y. Bar-Ness, "OFDM systems in the presence of phase noise: consequences and solutions," *IEEE Transactions on Communications*, vol. 52, pp. 1988–1996, Nov. 2004.
- [11] D. Petrovic, W. Rave, and G. Fettweis, "Effects of phase noise on OFDM systems with and without PLL: Characterization and compensation," *IEEE Transactions on Communications*, vol. 55, pp. 1607–1616, Aug. 2007.
- [12] H. Mehrpouyan, A. Nasir, S. Blostein, T. Eriksson, G. Karagiannidis, and T. Svensson, "Joint Estimation of Channel and Oscillator Phase Noise in MIMO Systems," *IEEE Transactions on Signal Processing*, vol. 60, pp. 4790–4807, Sep. 2012.
- [13] A. Lapidoth, "On phase noise channels at high SNR," in *Proceedings of the 2002 IEEE Information Theory Workshop*, pp. 1–4, Oct. 2002.
- [14] G. Durisi, A. Tarable, C. Camarda, R. Devassy, and G. Montorsi, "Capacity bounds for MIMO microwave backhaul links affected by phase noise," *IEEE Transactions on Communications*, vol. 62, pp. 920–929, Mar. 2014.
- [15] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Transactions on Information Theory*, vol. 60, pp. 7112–7139, Nov. 2014.
- [16] A. Pitarokoilis, S. K. Mohammed, and E. G. Larsson, "Effect of oscillator phase noise on the uplink performance of large MU-MIMO systems," in *50th Allerton Conference on Communication Control and Computing*, pp. 1190–1197, Oct. 2012.
- [17] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Linear Pre-Coding Performance in Measured Very-Large MIMO Channels," in *Vehicular Technology Conference (VTC Fall), 2011 IEEE*, pp. 1–5, Sept. 2011.
- [18] A. Demir, A. Mehrotra, and J. Roychowdhury, "Phase noise in oscillators: a unifying theory and numerical methods for characterization,"

- IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 47, pp. 655–674, May 2000.
- [19] B. Hassibi and B. Hochwald, “How much training is needed in multiple-antenna wireless links?,” *IEEE Transactions on Information Theory*, vol. 49, pp. 951 – 963, Apr. 2003.
- [20] T. L. Marzetta, “How much training is required for multiuser MIMO?,” in *Fortieth Asilomar Conference on Signals, Systems and Computers, 2006. ACSSC '06.*, pp. 359 –363, Nov. 2006.
- [21] T. Höhne and V. Ranki, “Phase noise in beamforming,” *IEEE Transactions on Wireless Communications*, vol. 9, pp. 3682 – 3689, Dec. 2010.
- [22] H. Yang and T. L. Marzetta, “Total energy efficiency of cellular large-scale antenna system multiple access mobile networks,” in *Proc. IEEE Online Conference on Green Communications*, pp. 27–32, Oct. 2013.
- [23] D. Petrovic, W. Rave, and G. Fettweis, “Common phase error due to phase noise in OFDM-estimation and suppression,” in *15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), 2004*, vol. 3, pp. 1901 – 1905, Sept. 2004.
- [24] A. Pitarokoilis, S. K. Mohammed, and E. G. Larsson, “On the optimality of single-carrier transmission in large-scale antenna systems,” *IEEE Wireless Communications Letters*, vol. 1, pp. 276 –279, Aug. 2012.

PAPER C

Achievable Rates of ZF Receivers in Massive MIMO with Phase Noise Impairments

Refereed article published in the Proc. of the Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, 2013.

©2013 IEEE. The layout has been revised and the bibliography has been updated.

Achievable Rates of ZF Receivers in Massive MIMO with Phase Noise Impairments

Antonios Pitarokoilis, Saif Khan Mohammed, and Erik G. Larsson

Abstract

The effect of oscillator phase noise on the sum-rate performance of large multi-user multiple-input multiple-output (MU-MIMO) systems is studied. A Rayleigh fading MU-MIMO uplink channel is considered, where channel state information (CSI) is acquired via training. The base station (BS), which is equipped with an excess of antenna elements, M , uses the channel estimate to perform zero-forcing (ZF) detection. A lower bound on the sum-rate performance is derived. It is shown that the proposed receiver structure exhibits an $O(\sqrt{M})$ array power gain. Additionally, the proposed receiver is compared with earlier studies that employ maximum ratio combining and it is shown that it can provide significant sum-rate performance gains at the medium and high signal-to-noise-ratio (SNR) regime. Further, the expression of the achievable sum rate provides new insights on the effect of various parameters on the overall system performance.

1 Introduction

In the last decades there has been an unprecedented increase in demand for high data rates over wireless cellular systems. Multiple-input multiple-output (MIMO) systems have attracted significant attention since they promise substantial gains in the capacity of wireless networks [1,2]. Such systems have gained popularity and are already part of modern cellular communication standards. The demand for increased high speed wireless access is expected to accelerate in the years to come. In addition, power consumption at the base stations (BS) of cellular systems is a major problem. Therefore, future cellular systems are required to exhibit increased spectral efficiency as well as improved energy efficiency.

Recently, it has been shown that using an excess of BS antenna elements, M , (Massive MIMO) can boost data rates and simultaneously it can reduce the required transmit power with very simple linear signal processing [3,4]. In [5] the authors show that for the case of MU-MIMO uplink channel with linear receivers, one can reduce the total transmitted power proportionally to M (or \sqrt{M}) for the case of perfect (or imperfect) channel state information (CSI), without compromising the total throughput. Further studies also highlight the gains in energy efficiency that are offered by Massive MIMO [6,7].

In practice, communication systems inevitably suffer from transceiver impairments. One such impairment is phase noise. Phase noise is introduced by imperfect oscillators that convert the baseband signal to the passband and vice versa. Ideal oscillators output a sinusoid that is stable in amplitude, frequency and phase. However, due to imperfections of the circuitry of the oscillators the phase of the sinusoid drifts randomly. Consequently, phase noise can significantly degrade the performance of coherent communication. In a system where CSI is acquired via uplink training, phase noise causes a partial loss of coherency between the estimated channel coefficients and the channel realizations during detection. The large array gain offered by Massive MIMO is a consequence of the coherent combining of received signals using the estimated channel gains. Therefore, the study of the effect of phase noise impairment on the performance of Massive MIMO systems is particularly important.

In earlier work [8], we have studied the effect of oscillator phase noise in the frequency selective Massive MIMO uplink when detection is performed

via time-reversal maximum-ratio-combining (TR-MRC). TR-MRC is optimal at low signal-to-noise-ratio (SNR), where the system performance is limited by thermal noise. However, in the high-SNR regime the system performance is interference limited and the MRC sum-rate performance saturates. Zero-forcing (ZF) receivers can provide improved performance in this regime since they effectively suppress multi-user interference. In this work we study the sum-rate performance of ZF receivers with phase noise impairments and compare it with the sum-rate performance of MRC receivers. To the best of authors' knowledge, this paper is the first to report such a study.

We consider a Rayleigh fading channel with K non-cooperative users and M BS antennas. Phase noise is introduced both at the transmitter and receiver side and is modelled as a Wiener phase noise process. The BS estimates the channel via uplink training and uses the acquired CSI to perform ZF detection. We provide an analytical formula for a lower bound on the sum-capacity and a coding strategy that achieves this lower bound. We prove that, just like MRC, ZF also offers an $O(\sqrt{M})$ array power gain. We show that ZF is superior to MRC in term of sum-rate performance in the high-SNR regime. For a desired per-user information rate, we also study the extra per-user transmit power required by MRC receivers when compared to the ZF receiver. The derived achievable sum-rate can provide new insights in the design of Massive MIMO systems with phase noise impairments.

2 System Model

We consider a frequency flat multi-user multiple-input multiple-output (MU-MIMO) uplink channel, with K single antenna users and M base station (BS) antenna elements. The channel between user k and antenna element m is given by $\sqrt{d_k}h_{m,k}$, where d_k and $h_{m,k}$ models the slow and fast fading components, respectively. In this work, we assume a block fading model, where $h_{m,k}$ is fixed for a block of $K + N_D$ channel uses and varies independently from one block to another. Since d_k changes at a much slower rate compared to $h_{m,k}$, we assume that it is fixed and known throughout the communication. The fast fading coefficients $h_{m,k}$ are assumed to be independent identically distributed (i.i.d.) circularly symmetric complex normal $\mathcal{CN}(0, 1)$ random variables.

Phase noise is introduced at the transmitter during up-conversion, when the baseband signal is multiplied with the carrier generated by the oscillator. The phase of the generated carrier drifts randomly, resulting in phase distortion of the transmitted signal. A similar phenomenon also happens at the receiver side during down-conversion of the bandpass signal to baseband. In the following, θ_k , $k = 1, \dots, K$ denotes the phase noise process at the k -th user and ϕ denotes the phase noise process at the M BS receivers. The latter implies identical phase noise processes at the BS antenna elements, i.e. we assume full coherency between the M BS receivers. This models the practical scenario of a centralized BS with a single oscillator output feeding the down-conversion module in each receiver. We further assume that the phase noise processes ϕ , θ_k , $k = 1, \dots, K$ are mutually independent. In this study every phase noise process is modelled as an independent Wiener process, which is a well-established model [9,10]. Therefore, the discrete time phase noise process at the BS antennas at time n is given by¹ $\phi[i] = \phi[i-1] + w[i]$, where $w[i] \sim \mathcal{N}(0, \sigma_\phi^2)$, $\sigma_\phi^2 \triangleq 4\pi^2 f_c^2 c T_s$ are independent identically distributed zero-mean Gaussian increments. f_c is the carrier frequency, T_s is the symbol interval and c is a constant that depends on the oscillator. Similarly, we can define the phase noise processes at the K users.

Let $x_k[i]$ be the symbol transmitted from the k -th user at time i . The received signal at the m -th BS antenna element at time i is then given by

$$y_m[i] = \sqrt{P} \sum_{k=1}^K \sqrt{d_k} h_{m,k} e^{-j(\phi[i] - \theta_k[i])} x_k[i] + n_m[i], \quad (1)$$

where $n_m[i] \sim \mathcal{CN}(0, \sigma^2)$ is additive white Gaussian noise (AWGN). Each user transmits a stream of i.i.d. $\mathcal{CN}(0, 1)$ information symbols (i.e. $x_k[i] \sim \mathcal{CN}(0, 1)$), that are independent of the information symbols of the other users. P denotes the average uplink transmitted power from each user.

3 Transmission Scheme and ZF Receiver

Motivated by the need for low-complexity channel estimation and detection algorithms, we propose the following block based uplink transmission

¹The discrete-time phase noise model is used since we are interested in the discrete-time complex baseband representation of the transmit and receive signals.

scheme. In the proposed scheme, a transmission block of $K + N_D$ channel uses consists of K channel uses (for uplink channel estimation) followed by the data phase (for data transmission) of duration N_D channel uses. For coherent demodulation, the BS needs to estimate the uplink channel. This is facilitated through the transmission of uplink pilot symbols during the training phase of each transmission block. The users transmit uplink training signals sequentially in time, i.e. at any given time only one user is transmitting uplink training signals and all other users are idle. To be precise, the k -th user sends an impulse signal of amplitude $\sqrt{P_p K}$ at the $(k - 1)$ -th channel use and is idle for the remaining portion of the training phase. Here, P_p is the average power transmitted by a user during the training phase. Therefore, using (1) the signal received at the m -th BS receiver at time $k - 1$, $k = 1, \dots, K$ is given by

$$y_m[k - 1] = \sqrt{P_p K d_k} h_{m,k} e^{-j(\phi[k-1] - \theta_k[k-1])} + n_m[k - 1]. \quad (2)$$

3.1 LMMSE Channel Estimation

Based on the received signal during training, the effective MIMO channel is estimated in order to enable coherent detection of the received symbols. The received symbols during training, given by (2), can be expressed in the following matrix-vector form

$$\mathbf{Y} = \sqrt{P_p K} \mathbf{H} \mathbf{\Omega} \mathbf{D}^{1/2} + \mathbf{N} \quad (3)$$

where $[\mathbf{Y}]_{m,k} = y_m[k - 1]$, $[\mathbf{H}]_{m,k} = h_{m,k}$, $\mathbf{\Omega} = \text{diag}\{e^{-j(\phi[0] - \theta_1[0])}, \dots, e^{-j(\phi[K-1] - \theta_K[K-1])}\}$, $\mathbf{D} = \text{diag}\{d_1, \dots, d_K\}$, $d_k > 0$, $\forall k = 1, \dots, K$ and $[\mathbf{N}]_{m,k} = n_m[k - 1]$. Also, observe that $\mathbf{\Omega} \mathbf{\Omega}^H = \mathbf{\Omega}^H \mathbf{\Omega} = \mathbf{I}_K$. We define the effective MIMO channel matrix as $\mathbf{G} \triangleq \mathbf{H} \mathbf{\Omega}$. The LMMSE estimate $\hat{\mathbf{G}}$ for \mathbf{G} based on (3) is given by [11]

$$\hat{\mathbf{G}} = \mathbf{Y} \text{diag} \left\{ \frac{\sqrt{P_p K d_1}}{P_p K d_1 + \sigma^2}, \dots, \frac{\sqrt{P_p K d_K}}{P_p K d_K + \sigma^2} \right\}. \quad (4)$$

As expected, the channel estimate is distorted by the AWGN and by the phase noise at the transmitter and at the BS. In the following, we treat this channel estimate as the true channel and perform equalization based on that. Conditioned on Ω , it is clear from (3) that the entries of \mathbf{Y} are Gaussian distributed. Since $\mathbf{G} = \mathbf{H}\Omega$ and $\hat{\mathbf{G}}$ is a column wise scaled version of \mathbf{Y} (see (4)), it follows that conditioned on Ω , \mathbf{G} , $\hat{\mathbf{G}}$ and $\mathcal{E} \triangleq \hat{\mathbf{G}} - \mathbf{G}$ are all Gaussian distributed. Therefore, conditioned on Ω , the LMMSE estimate in (4) is actually the MMSE estimate and therefore $\hat{\mathbf{G}}$ and \mathcal{E} are independent (conditioned on Ω). Since we are interested in evaluating the performance of ZF receivers, the equalizer would simply be the pseudo-inverse of $\hat{\mathbf{G}}$, i.e. $\hat{\mathbf{G}}^\dagger \triangleq (\hat{\mathbf{G}}^H \hat{\mathbf{G}})^{-1} \hat{\mathbf{G}}^H$. Since $\hat{\mathbf{G}}$ and \mathcal{E} are independent (conditioned on Ω), it therefore follows that

$$\mathbb{E}_{\hat{\mathbf{G}}, \mathcal{E} | \Omega} [\hat{\mathbf{G}}^\dagger \mathcal{E}] = \mathbb{E}_{\hat{\mathbf{G}} | \Omega} [\hat{\mathbf{G}}^\dagger] \mathbb{E}_{\mathcal{E} | \Omega} [\mathcal{E}] = 0 \quad (5)$$

since $\mathbb{E}_{\mathcal{E} | \Omega} [\mathcal{E}] = 0$.

3.2 Zero-Forcing (ZF) Equalization

During the data interval, the received symbol at time $i \geq K$ at the m -th BS antenna is given by

$$y_m[i] = \sqrt{P_D} \sum_{k=1}^K h_{m,k} e^{-j(\phi[i] - \theta_k[i])} \sqrt{d_k} x_k[i] + n_m[i]. \quad (6)$$

Therefore, the matrix-vector form of the channel input-output relation is given by

$$\mathbf{y}[i] = \sqrt{P_D} \mathbf{H} \Phi[i] \mathbf{D}^{1/2} \mathbf{x}[i] + \mathbf{n}[i], \quad (7)$$

where $\Phi[i] \triangleq \text{diag} \{e^{-j(\phi[i] - \theta_1[i])}, \dots, e^{-j(\phi[i] - \theta_K[i])}\}$ and P_D is the per-user average transmit power constraint during the data interval. In general, the relation between P_D and P_p can be arbitrary. However, in this work we restrict to the case where $P_p = \beta P_D$, $\beta > 0$.

Since $\mathbf{G} = \mathbf{H}\mathbf{\Omega}$ and $\mathbf{\Omega}\mathbf{\Omega}^H = \mathbf{I}_K$, it follows that $\mathbf{H} = \mathbf{G}\mathbf{\Omega}^H$. Also, since $\mathbf{G} = \hat{\mathbf{G}} - \mathbf{\mathcal{E}}$, we have $\mathbf{H} = \mathbf{G}\mathbf{\Omega}^H = (\hat{\mathbf{G}} - \mathbf{\mathcal{E}})\mathbf{\Omega}^H$. Therefore, replacing \mathbf{H} by $(\hat{\mathbf{G}} - \mathbf{\mathcal{E}})\mathbf{\Omega}^H$ in (7) we get

$$\mathbf{y}[i] = \sqrt{P_D} (\hat{\mathbf{G}} - \mathbf{\mathcal{E}}) \mathbf{\Omega}^H \mathbf{\Phi}[i] \mathbf{D}^{1/2} \mathbf{x}[i] + \mathbf{n}[i]. \quad (8)$$

Then, the detected symbol vector $\hat{\mathbf{x}}[i]$ at time i is given by

$$\begin{aligned} \hat{\mathbf{x}}[i] &= \hat{\mathbf{G}}^\dagger \mathbf{y}[i] = \sqrt{P_D} \hat{\mathbf{G}}^\dagger \mathbf{\Omega}^H \mathbf{\Phi}[i] \mathbf{D}^{1/2} \mathbf{x}[i] \\ &\quad - \sqrt{P_D} \hat{\mathbf{G}}^\dagger \mathbf{\mathcal{E}} \mathbf{\Omega}^H \mathbf{\Phi}[i] \mathbf{D}^{1/2} \mathbf{x}[i] + \hat{\mathbf{G}}^\dagger \mathbf{n}[i], \end{aligned} \quad (9)$$

where $\hat{\mathbf{G}}^\dagger \triangleq (\hat{\mathbf{G}}^H \hat{\mathbf{G}})^{-1} \hat{\mathbf{G}}^H$ is the pseudo-inverse of $\hat{\mathbf{G}}$.

4 Achievable Rates

In this work we are interested in evaluating the sum-rate performance for the system under study. In the following, we describe the approach to derive an achievable sum rate for the Massive MIMO uplink with phase noise impairments when ZF reception is employed. From (9) the detected symbol for the k -th user at time i is given by

$$\begin{aligned} \hat{x}_k[i] &= \sqrt{P_D} d_k e^{j(\phi[k-1] - \theta_k[k-1])} e^{-j(\phi[i] - \theta_k[i])} x_k[i] \\ &\quad - \sqrt{P_D} e_k^T \hat{\mathbf{G}}^\dagger \mathbf{\mathcal{E}} \mathbf{\Omega}^H \mathbf{\Phi}[i] \mathbf{D}^{1/2} \mathbf{x}[i] + e_k^T \hat{\mathbf{G}}^\dagger \mathbf{n}[i], \end{aligned} \quad (10)$$

where e_k is the K -dimensional all-zero column vector that has a single 1 at position k . Let $A_k[i] \triangleq \sqrt{P_D} d_k e^{j(\phi[k-1] - \theta_k[k-1])} e^{-j(\phi[i] - \theta_k[i])}$. Since the exact phase noise realizations are unknown, the desired symbol $x_k[i]$ is rotated by an unknown phase. Further, the statistics of $A_k[i]$ depends on i . However, the statistics of the phase noise processes is known to the BS and as a result it knows the mean value of $A_k[i]$. In (10) we add and subtract the term $\mathbb{E}[A_k[i]] x_k[i]$. The added term $\mathbb{E}[A_k[i]] x_k[i]$ is the desired signal term and the term $(A_k[i] - \mathbb{E}[A_k[i]]) x_k[i]$ is relegated to an effective noise term. Therefore, (10) can be expressed as

$$\hat{x}_k[i] = \mathbb{E}[A_k[i]] x_k[i] + \text{EN}_k[i] \quad (11)$$

where

$$\begin{aligned}
 \text{EN}_k[i] &\triangleq \text{IF}_k[i] + \text{MUI}_k[i] + \text{AN}_k[i] \\
 \text{IF}_k[i] &\triangleq (A_k[i] - \mathbb{E}[A_k[i]]) x_k[i], \\
 \text{MUI}_k[i] &\triangleq -\sqrt{P_D} e_k^T \hat{\mathbf{G}}^\dagger \boldsymbol{\varepsilon} \boldsymbol{\Omega}^H \boldsymbol{\Phi}[i] \mathbf{D}^{1/2} \mathbf{x}[i], \\
 \text{AN}_k[i] &\triangleq e_k^T \hat{\mathbf{G}}^\dagger \mathbf{n}[i].
 \end{aligned}$$

Proposition 1. *In (11) the desired signal term $\mathbb{E}[A_k[i]] x_k[i]$ and the effective additive noise term $\text{EN}_k[i]$ are uncorrelated.*

Proof. It is straightforward to see that $\mathbb{E}[\mathbb{E}[A_k[i]] x_k[i] \text{IF}_k^*[i]] = 0$. Also $\mathbb{E}[\mathbb{E}[A_k[i]] x_k[i] \text{AN}_k^*[i]] = 0$ by the assumption that $\mathbf{x}[i]$ and $\mathbf{n}[i]$ are uncorrelated. Finally,

$$\begin{aligned}
 \mathbb{E}[\text{MUI}_k[i] x_k^*[i] \mathbb{E}[A_k^*[i]]] &= -\sqrt{P_D} \mathbb{E} \left[e_k^T \hat{\mathbf{G}}^\dagger \boldsymbol{\varepsilon} \boldsymbol{\Omega}^H \boldsymbol{\Phi}[i] \mathbf{D}^{1/2} e_k \right] \\
 &\cdot \mathbb{E}[A_k^*[i]] = -\sqrt{P_D d_k} \mathbb{E}_\Omega \left[e_k^T \mathbb{E}_{\hat{\mathbf{G}}, \boldsymbol{\varepsilon} | \Omega} \left[\hat{\mathbf{G}}^\dagger \boldsymbol{\varepsilon} \right] e_k e^{-j(\phi[i] - \theta_k[i])} \right. \\
 &\cdot \left. e^{j(\phi[k-1] - \theta_k[k-1])} \right] \mathbb{E}[A_k^*[i]] = 0,
 \end{aligned}$$

which follows since $\mathbb{E}_{\hat{\mathbf{G}}, \boldsymbol{\varepsilon} | \Omega} \left[\hat{\mathbf{G}}^\dagger \boldsymbol{\varepsilon} \right] = 0$ from (5). Hence, we can conclude that $\mathbb{E}[\mathbb{E}[A_k[i]] x_k[i] \text{EN}_k^*[i]] = 0$. \square

We also provide a result that will be proved useful in the derivation of the achievable rates.

Theorem 1. *The mean value of $A_k[i]$ and the variance $\text{Var}(\text{EN}_k[i]) \triangleq \mathbb{E}[|\text{EN}_k[i] - \mathbb{E}[\text{EN}_k[i]]|^2]$ are given by*

$$\mathbb{E}[A_k[i]] = \sqrt{P_D d_k} \kappa_k[i], \quad (12)$$

$$\text{Var}(\text{EN}_k[i]) = P_D d_k (1 - \kappa_k^2[i]) + \sigma^2 C_k, \quad (13)$$

where $\kappa_k[i] = e^{-\sigma_\phi^2(i-(k-1))}$ and

$$C_k \triangleq \left(\sum_{l=1}^K \frac{\frac{P_D}{\sigma^2} d_l}{\beta \frac{P_D}{\sigma^2} K d_l + 1} + 1 \right) \frac{\beta \frac{P_D}{\sigma^2} K d_k + 1}{(M - K) \beta \frac{P_D}{\sigma^2} K d_k}.$$

In the following, we describe a method to derive an achievable information rate for the k -th user. Similar techniques have been used earlier in [8, 12, 13]. Observe that, in a given coherence interval, $\mathbb{E}[A_k[i]]$ and $\text{Var}(\text{EN}_k[i])$ are different for different i . However $\mathbb{E}[A_k[i]]$ and $\text{Var}(\text{EN}_k[i])$ for a fixed k and i are the same across multiple coherence intervals. Further, the realizations of $\text{EN}_k[i]$ for a fixed i and k are i.i.d. across multiple coherence intervals. Therefore, for the k -th user, we propose to encode information using N_D different channel codes, one channel code for each i . That is, in the i -th channel code information is coded across the i -th channel use of each coherence interval.

We proceed by describing the derivation of a lower bound for the rate of the i -th channel code for the k -th user. Since the BS has knowledge of the statistics of the channel coefficients and the phase noise processes but not of the exact realization, the term $\mathbb{E}[A_k[i]]$ can be computed and is a constant depending on the user k and the channel use i during the data transmission interval. On the other hand, the exact probability distribution of the effective noise term is complicated to compute. However, since the input symbols are uncorrelated to the effective noise $\text{EN}_k[i]$ (from Proposition 1), a lower bound on the achievable information rate can be derived by considering the worst case uncorrelated noise having the same variance as $\text{EN}_k[i]$. Given that the input symbols are Gaussian, the worst case uncorrelated additive noise is also zero mean Gaussian with variance equal to the variance of $\text{EN}_k[i]$. Therefore the mutual information between the detected symbol $\hat{x}_k[i]$ and the transmitted symbol is lower bounded by

$$I(\hat{x}_k[i]; x_k[i]) \geq \log_2 \left(1 + \frac{(\mathbb{E}[A_k[i]])^2}{\text{Var}(\text{EN}_k[i])} \right). \quad (14)$$

Theorem 2. *An achievable rate for the i -th channel code of the k -th user is given by*

$$R_k^{ZF}[i] = \log_2 \left(1 + \frac{\frac{P_D}{\sigma^2} d_k \kappa_k^2[i]}{\frac{P_D}{\sigma^2} d_k (1 - \kappa_k^2[i]) + C_k} \right) \quad (15)$$

Proof. The result follows directly by substituting (12) and (13) into (14). \square

The information rate for the no-phase noise case (i.e. $c = 0$, perfect oscilla-

tor) can be derived by substituting $\kappa_k[i] = 1$ in 15, and is given by

$$\mathcal{R}_k = \log_2 \left(1 + \frac{\beta K(M-K) \left(\frac{P_D}{\sigma^2}\right)^2 d_k^2}{\left(\beta \frac{P_D}{\sigma^2} K d_k + 1\right) \left(\sum_{l=1}^K \frac{\frac{P_D}{\sigma^2} d_l}{\beta \frac{P_D}{\sigma^2} K d_{l+1}} + 1\right)} \right). \quad (16)$$

For the sake of comparison, we also derive the information rate expression for maximum-ratio-combining (MRC) when LMMSE channel estimation is used (using similar arguments to the ZF case). The i -th code of user k achieves the rate

$$R_k^{\text{MRC}}[i] = \log_2 \left(1 + \frac{\frac{P_D}{\sigma^2} d_k \kappa_k^2[i]}{\frac{P_D}{\sigma^2} d_k (1 - \kappa_k^2[i]) + C_k^{\text{MRC}}} \right), \quad (17)$$

where $C_k^{\text{MRC}} \triangleq \frac{\beta P_D}{\sigma^2} K d_k + 1 \left(\frac{P_D}{\sigma^2} \sum_{l=1}^K d_l + 1 \right)$. We conclude the section by defining the average achievable sum-rate for ZF as

$$R^{\text{ZF}} \triangleq \frac{1}{N_D + K} \sum_{k=1}^K \sum_{i=K}^{N_D+K-1} R_k^{\text{ZF}}[i], \quad (18)$$

where the normalization factor $\frac{1}{N_D+K}$ accounts for the loss of spectral efficiency due to training. Similarly, one can define the average achievable sum-rate for the MRC and no-phase noise cases.

5 Results - Discussion

In this section we use the expressions derived in (15), (16), (17) and (18) to present our main results. In the plots we always consider $\mathbf{D} = \mathbf{I}_K$, which corresponds to the case of equal path loss for all the users, whereas in the theoretical derivations we consider an arbitrary, known and deterministic slow fading matrix \mathbf{D} . We start with a proposition on the information rate achievable at high SNR.

Proposition 2. *In the high-SNR regime the rate that user k can achieve at the i -th channel use saturates to the value*

$$\lim_{\frac{P_D}{\sigma^2} \rightarrow \infty} R_k^{\text{ZF}}[i] \rightarrow \log_2 \left(1 + \frac{\kappa_k^2[i]}{1 - \kappa_k^2[i]} \right) \quad (19)$$

for ZF and at the value

$$\lim_{\frac{P_D}{\sigma^2} \rightarrow \infty} R_k^{\text{MRC}}[i] \rightarrow \log_2 \left(1 + \frac{d_k \kappa_k^2[i]}{d_k (1 - \kappa_k^2[i]) + \frac{1}{M} \sum_{l=1}^K d_l} \right) \quad (20)$$

for MRC.

A desirable property of Massive MIMO is the array power gain they offer, facilitating the design of power efficient communication systems. In earlier work [5,8] it has been shown that in the case of Massive MIMO with imperfect CSI, one can scale down the total transmit power by \sqrt{M} as the number of BS antennas M increase, while maintaining a fixed positive desired information rate to each user. In this work, we extend the result to the case of Massive MIMO uplink impaired with phase noise when ZF receivers are employed.

Proposition 3. *An $O(\sqrt{M})$ array power gain is achievable. This implies that a fixed non-zero per user information rate can be achieved if the total transmitted power is reduced by 1.5 dB and at the same time the number of BS antenna elements is doubled.*

Proof. From (15) and by the substitution $P_D = \frac{E_u}{M^\alpha}$, $\alpha > 0$ we have that

$$R_k^{\text{ZF}}[i] = \log_2 \left(1 + \frac{\frac{E_u}{\sigma^2} d_k \kappa_k^2[i]}{\frac{E_u}{\sigma^2} d_k (1 - \kappa_k^2[i]) + M^\alpha C_k} \right).$$

We look for the largest $\alpha > 0$ such that $\lim_{M \rightarrow \infty} R_k[i] \rightarrow r > 0 \Leftrightarrow \lim_{M \rightarrow \infty} M^\alpha C_k \rightarrow \text{const.}$

$$\begin{aligned} M^\alpha C_k &= M^\alpha \left(\sum_{l=1}^K \frac{\frac{E_u}{\sigma^2} d_l}{\beta \frac{E_u}{\sigma^2} K d_l + 1} \right) \frac{\beta K \frac{E_u}{\sigma^2} d_k + 1}{(M - K) \beta \frac{E_u}{\sigma^2} K d_k} \\ &+ M^\alpha \frac{\beta K \frac{E_u}{\sigma^2} d_k + 1}{(M - K) \beta \frac{E_u}{\sigma^2} K d_k} \\ &= \left(\sum_{l=1}^K \frac{\frac{E_u}{\sigma^2} d_l M^\alpha}{\beta \frac{E_u}{\sigma^2} K d_l + M^\alpha} \right) \frac{\beta K \frac{E_u}{\sigma^2} d_k + M^\alpha}{(M - K) \beta \frac{E_u}{\sigma^2} K d_k} + \frac{\beta K \frac{E_u}{\sigma^2} d_k + M^\alpha}{\frac{(M-K)}{M^\alpha} \beta \frac{E_u}{\sigma^2} K d_k} \end{aligned}$$

As $M \rightarrow \infty$, the first term converges to a finite positive constant if and only $\alpha \leq 1$, whereas the second term converges to a finite positive constant if and

only if $\alpha \leq 1/2$. For $\alpha > 1/2$, the second term is unbounded. Therefore, as $M \rightarrow \infty$, with $P_D = \frac{E_u}{M^\alpha}$, we have

$$\lim_{M \rightarrow \infty} R_k^{\text{ZF}}[i] \rightarrow \begin{cases} \log_2 \left(1 + \frac{\frac{E_u}{\sigma^2} d_k \kappa_k^2 [i]}{\frac{E_u}{\sigma^2} d_k (1 - \kappa_k^2 [i]) + \frac{1}{\beta \frac{E_u}{\sigma^2} K d_k}} \right), & \alpha = 1/2 \\ \log_2 \left(1 + \frac{\kappa_k^2 [i]}{1 - \kappa_k^2 [i]} \right), & \alpha < 1/2 \\ 0, & \alpha > 1/2 \end{cases}$$

□

In Fig. 1 we plot the minimum required $\frac{P_D}{\sigma^2}$ to achieve a fixed desired per-user information rate $r = 1.5$ bpcu as a function of the number of BS antennas M for fixed $K = 10$ users, $N_D = 100$ channel uses, $T_s = 10^{-6}$ s and $c = 4.7 \times 10^{-18}$ (rad Hz) $^{-1}$. We plot the curves for ZF and MRC with and without phase noise. It is clear that at large M , when the number of BS antennas doubles the minimum required power decreases by 1.5 dB. This verifies the $O(\sqrt{M})$ array power gain as stated in Proposition 3.

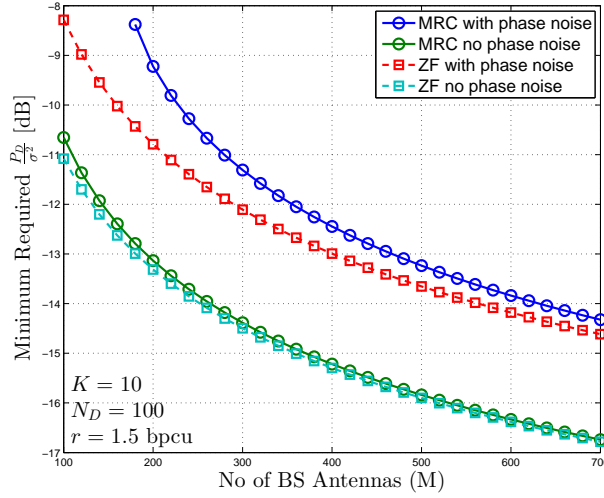


Figure 1: Minimum required $\frac{P_D}{\sigma^2}$ to achieve a fixed per-user information rate of $r = 1.5$ bpcu as a function of increasing M for fixed $K = 10$ users, $N_D = 100$ channel uses, $T_s = 10^{-6}$ s and $c = 4.7 \times 10^{-18}$ (rad Hz) $^{-1}$.

Further, the superior performance of the ZF over MRC in the high-SNR regime is clear. When it is desired to achieve a per-user rate of $r = 1.5$

bpcu with $\frac{P_D}{\sigma^2} = -10$ dB, ZF with phase noise requires the use of about $M = 160$ BS antennas. However, the MRC system with phase noise for the same parameters requires approximately 80 additional BS antennas. This motivates us to study the $\frac{P_D}{\sigma^2}$ gap between ZF and MRC reception. In Fig. 2 we plot this gap as a function of the desired per-user information rate for fixed $M = 200$ BS antennas and $N_D = 150$ channel uses for various values of the number of users, K . It is clear that there is a significant gain in the minimum required power for per-user information rates larger than 1 bpcu. The gain is more pronounced as the number of users increases. The reason for this behaviour is that for $r > 1$ bpcu and for K sufficiently large, MRC is limited by interference, whereas the ZF receiver effectively suppresses multi-user interference yielding better performance. At a per-user rate of $r \approx 0.9$ bpcu there is a turning point. Below this value we operate in the low spectral efficiency regime and the system is limited by the thermal noise. As a result MRC exhibits better performance whereas ZF underutilizes the available degrees of freedom.

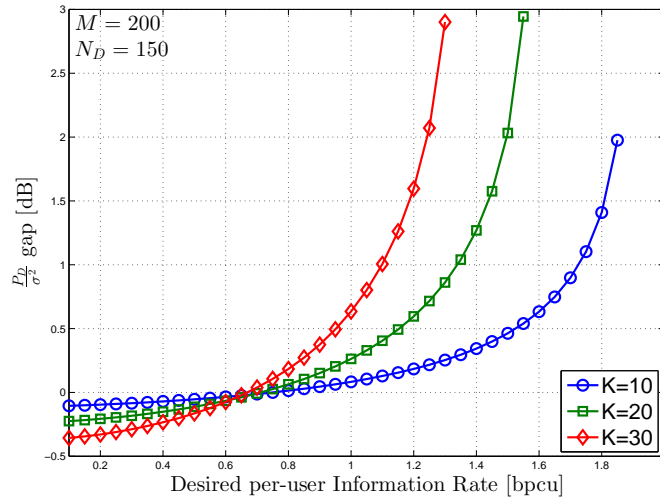


Figure 2: $\frac{P_D}{\sigma^2}$ gap between ZF and MRC in [dB] to achieve a desired per-user information rate [bpcu], for fixed $M = 200$ BS antennas, $N_D = 150$ channel uses, $T_s = 10^{-6}$ s and $c = 4.7 \times 10^{-18}$ (rad Hz) $^{-1}$ for various values of the number of users, K .

In conclusion, we studied the effect of oscillator phase noise in a frequency flat Massive MIMO uplink channel when ZF equalization is employed at

the BS. We provided an analytical expression on the achievable sum-rate and we described a coding strategy that achieves the derived sum-rate. Based on the derived expression we proved that an $O(\sqrt{M})$ array power gain is still attainable. Further, we showed that even in the presence of phase noise, ZF equalization performs significantly better than MRC at high spectral efficiencies. The derived achievable sum-rate expressions can provide additional insight on the effect of oscillator phase noise in Massive MIMO systems with ZF equalization.

Bibliography

- [1] G. Foschini and M. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, Mar. 1998.
- [2] D. Gesbert, M. Kountouris, R. W. Heath Jr., C.-B. Chae, and T. Sälzer, "Shifting the MIMO paradigm," *IEEE Signal Processing Magazine*, vol. 24, pp. 36–46, Sept. 2007.
- [3] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
- [4] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, pp. 40–60, Jan. 2013.
- [5] H. Q. Ngo, E. Larsson, and T. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Transactions on Communications*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.
- [6] S. K. Mohammed and E. G. Larsson, "Single-User Beamforming in Large-Scale MISO Systems with Per-Antenna Constant-Envelope Constraints: The Doughnut Channel," *IEEE Transactions on Wireless Communications*, vol. 11, no. 11, pp. 3992–4005, Nov. 2012.
- [7] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of Cellular Networks: How Many Antennas Do We Need?," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

-
- [8] A. Pitarokoilis, S. K. Mohammed, and E. G. Larsson, "Effect of oscillator phase noise on the uplink performance of large MU-MIMO systems," in *Proc. of the 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2012, Monticello, IL, pp. 1190–1197, Oct. 2012.
- [9] A. Demir, A. Mehrotra, and J. Roychowdhury, "Phase noise in oscillators: a unifying theory and numerical methods for characterization," *IEEE Trans. on Circuits and Systems I: Fundamental Theory and Applications*, vol. 47, no. 5, pp. 655–674, May 2000.
- [10] D. Petrovic, W. Rave, and G. Fettweis, "Effects of phase noise on OFDM systems with and without PLL: Characterization and compensation," *IEEE Transactions on Communications*, vol. 55, no. 8, pp. 1607–1616, Aug. 2007.
- [11] S. M. Kay, *Fundamentals of statistical signal processing: estimation theory*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.
- [12] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 951 – 963, Apr. 2003.
- [13] T. L. Marzetta, "How much training is required for multiuser MIMO?," in *Proc. 40th Asilomar Conference on Signals, Systems and Computers*, pp. 359–363, Nov. 2006.

PAPER D

ML Detection in Phase Noise Impaired SIMO Channels with Uplink Training

Refereed article published in the IEEE Transactions on Communications 2016. The version here contains one more Appendix (Appendix 8) and is publicly available at <http://arxiv.org/abs/1505.06553>.

©2016 IEEE. The layout has been revised.

ML Detection in Phase Noise Impaired SIMO Channels with Uplink Training

Antonios Pitarokoilis, Emil Björnson, and Erik G. Larsson

Abstract

The problem of maximum likelihood (ML) detection in training-assisted single-input multiple-output (SIMO) systems with phase noise impairments is studied for two different scenarios, i.e. the case when the channel is deterministic and known (constant channel) and the case when the channel is stochastic and unknown (fading channel). Further, two different operations with respect to the phase noise sources are considered, namely, the case of identical phase noise sources and the case of independent phase noise sources over the antennas. In all scenarios the optimal detector is derived for a very general parameterization of the phase noise distribution. Further, a high signal-to-noise-ratio (SNR) analysis is performed to show that symbol-error-rate (SER) floors appear in all cases. The SER floor in the case of identical phase noise sources (for both constant and fading channels) is independent of the number of antenna elements. In contrast, the SER floor in the case of independent phase noise sources is reduced when increasing the number of antenna elements (for both constant and fading channels). Finally, the system model is extended to multiple data channel uses and it is shown that the conclusions are valid for these setups, as well.

1 Introduction

The demand on wireless data services is expected to increase significantly over the next decade. Hence, next generation wireless networks must provide substantially larger data rates. Recently, it has been shown that massive multiple-input multiple-output (*Massive MIMO*) can provide substantial gains in spectral efficiency and radiated energy efficiency [1–3]. In Massive MIMO, K non-cooperative users are served by a base station (BS) with M BS antennas over the same time and frequency resources. When M is significantly larger than K (e.g., one order of magnitude) linear transmit and receive processing techniques are close to optimal and the minimum required radiated power can be reduced as a function of M when a fixed information rate is desired [4, 5].

In Massive MIMO, the BS uses estimated channel impulse responses to coherently combine the received uplink signals. The quality of the estimated channel state information (CSI) has direct impact on the performance of Massive MIMO systems. Hardware impairments further degrade the acquired channel knowledge. In addition, the deployment of Massive MIMO systems requires the use of inexpensive hardware, so that the monetary cost remains low. Such equipment is likely to have limited accuracy. Hence, the study of the impact of hardware impairments is of particular importance and relevance in Massive MIMO systems. Recently, this area has attracted significant research interest [6].

An unavoidable hardware impairment in wireless communications is phase noise. Phase noise is introduced in communications systems during the upconversion of the baseband signal to passband and vice versa due to imperfections in the circuitry of local oscillators. Ideally, the local oscillators should produce a sinusoidal wave that is perfectly stable in terms of amplitude, frequency and phase. In the frequency domain that would correspond to a Dirac impulse located at the carrier frequency. However, the phase of the generated carrier of realizable oscillators typically fluctuates. This is manifested by a spectral widening around the carrier frequency in the power spectral density of the local oscillator output. Phase noise can cause significant degradation in scenarios where it varies faster than the channel fading. This happens when the variance of the phase noise innovations and the coherence interval of the channel fading are large [7, Section IV.C]. Some scenarios where the phase noise degradation dominates over the degradation due to channel variation are fixed indoor communication

and Line-of-Sight (LoS) communication, such as WiFi at millimeter-wave frequencies and wireless broadband-to-home services, respectively. Further, phase noise causes a random rotation of the information signal, i.e. it is a multiplicative distortion. This makes the analysis and mitigation of the phase noise considerably more involved in comparison to additive distortions, such as quantization noise and generic non-linearities. In fact, it appears that Massive MIMO systems are less robust to phase noise than hardware impairments modeled as additive distortions [7].

The problem of calculating the capacity of phase noise impaired systems is particularly challenging. Closed-form expressions are not available even for the simplest cases. In [8] the author derives the first two terms of the high signal-to-noise-ratio (SNR) expansion of the capacity of a phase noise impaired non-fading single-input single-output (SISO) system for any phase noise process that is ergodic, stationary, and has finite differential entropy rate. In [9] the first two terms of the high-SNR capacity expansion for the block memoryless phase noise channel are derived. In [10] a high-SNR capacity upper bound for the Wiener phase noise MIMO channel is derived. Recently, the authors in [11] report approximate upper and lower bounds on the high-SNR capacity for the multiple-input single-output (MISO) and single-input multiple-output (SIMO) phase noise channels and compare the cases where separate and common oscillators are used. Lower bounds on the sum-capacity of multi-user Massive MIMO systems with linear reception and phase noise impairments are recently reported in [7], [12] and [13].

The problem of data detection in non-fading channels with phase noise impairments has been extensively studied in the literature. In [14] the optimal binary detector for partially coherent channels is derived. Detectors that are optimal in the high-SNR regime are derived in [15]. In [16] the problem of optimal symbol-by-symbol (SBS) detection in SISO systems is investigated when the carrier phase is unknown and it is shown that the computational complexity is prohibitive in the general case. A suboptimal but implementable algorithm is derived when the unknown carrier phase stays constant for a block of consecutive symbols and its performance is compared to the case of exact carrier phase information. In [17] a simulation-based phase noise model is used and the existence EVM floors for SISO systems is shown. The authors of [18] derive algorithms for SISO phase noise channels without fading based on factor graphs and the sum-product algorithm. An extension of this work for single-user MIMO systems is given in [19], where an estimate of the channel is inserted into the likelihood function as if this

estimate were equal to the true channel—resulting in a mismatched detector. A set of soft metrics for the single-user non-fading phase noise channel under various assumptions is derived in [20] and their performance is compared. In [21] an algorithm for joint data detection and phase noise estimation is derived for single-user MIMO systems and its performance is compared to a derived Cramér-Rao bound.

Even though the problem of phase noise in communication systems is extensively studied, there are still many open questions. In particular, the effect of phase noise on beamforming is not fully understood yet. In [22] the authors study the effect of phase noise in the *error vector magnitude* (EVM) of an antenna array. They show through analysis and measurements that the EVM at the direction of the main lobe is reduced when independent phase noise sources are used. In [12] achievable rates for the uplink transmission of Massive MIMO systems with time-reversal maximum-ratio-combining in frequency-selective channels are derived for the case of identical and independent phase noise sources. It is observed that the use of independent phase noise sources at the BS results in an increased achievable rate. This result is also supported by a toy example, where the actual capacity can be easily computed and shows that the capacity with independent phase noise sources is larger than the capacity with a single phase noise source. A similar result is reported in [7], where the authors show that the phase noise variance can be allowed to increase logarithmically with M when independent phase noise sources are used without losing much in performance. This is not true in the case of identical phase noise sources. Finally, the authors of [11] observe that the *phase noise number* (i.e., the second term in the high-SNR capacity expansion) is higher in the case of independent phase noise sources. The effect of phase noise in linear precoders in downlink Massive MIMO systems is studied recently in [23] and in [24]. In [24] the effect of imperfect hardware and the number of oscillators is investigated in distributed Massive MIMO downlink systems and it is shown that superior performance is achieved with independent phase noise sources.

The results in previous works are not conclusive in the sense that they involve lower or upper bounds on the capacity of the investigated systems. These bounds are useful as the calculation of the exact capacity is an exceedingly complicated problem, however, they do not provide us with a conclusive answer on the fundamental difference between the choice of a single or multiple local oscillators. Motivated by that, we rigorously derive the optimal detector in phase noise impaired SIMO systems with uplink training for various cases of interest. Our findings explain the effects noted

in the prior work; that is, the fact that superior performance was observed in the case of independent phase noise sources. This phenomenon becomes particularly apparent in the high-SNR regime and for a large number of antennas.

We consider a single-user single-input multiple-output (SU-SIMO) system with phase impairments at the multi-antenna receiver. The maximum likelihood (ML) detector is derived under two assumptions on the phase noise processes, namely, when identical (*synchronous* operation) and independent phase noise processes (*non-synchronous* operation) are assumed. The detectors are explicitly given under the assumptions of either constant or fading channels. The phase noise impairments are modeled so that most reasonable distributions of the phase noise increments can be treated in a unified framework. A high-SNR analysis is provided in all the examined scenarios and conclusions are drawn with respect to the symbol-error-rate (SER) performance of the detector when the number of receive antennas, M , increases. We observe that in the synchronous operation, an SER floor due to phase noise appears for both assumptions on the channel fading. The SER floor depends on the severity of the phase noise impairments and is independent of the number of receive antennas, M . An SER floor at high-SNR appears also in the non-synchronous cases. However, it is shown that this floor can be made arbitrarily small by increasing M .

2 System Model

In this paper, a single-antenna user communicates with a BS equipped with M antenna elements, which are impaired with phase noise. A simple transmission protocol is considered, which consists of two channel uses (see Fig. 1). In the first channel use a known symbol (pilot) is transmitted and in the second an unknown information symbol is transmitted. Two different cases are treated with respect to the knowledge of the wireless channel. Namely, in the first case, termed as *constant channel (CC)*, the channel is assumed deterministic and known at the receiver [25], [26]. Hence, the transmitted symbol is observed in the presence of only additive noise and multiplicative phase noise. In the second case, termed as *fading channel (FC)*, the wireless channel is, additionally, assumed unknown at the receiver and Rayleigh fading. We start with the description of the CC for simplicity and subsequently we describe the extension to the FC.

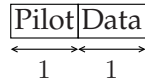


Figure 1: Two slot transmission protocol.

2.1 Constant Channel (CC)

During the first channel use, the received complex baseband symbol, x_m , at the m -th BS antenna is given by

$$x_m = \sqrt{\rho}g_m e^{j\theta_m} + w_m, \quad m = 1, \dots, M, \quad (1)$$

where w_m is the m -th component of the additive white Gaussian noise (AWGN) vector, \mathbf{w} , distributed as a circularly symmetric complex Gaussian random vector, $\mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$, θ_m is the unknown initial phase reference uniformly distributed in the interval $[-\pi, \pi)$. The positive scalar g_m is a known amplitude and ρ is the measured SNR at the m -th BS antenna when $g_m = 1$. In the second channel use the received symbol at the m -th BS antenna is given by

$$y_m = \sqrt{\rho}g_m e^{j(\theta_m + \phi_m)} s + z_m, \quad m = 1, \dots, M, \quad (2)$$

where z_m is the m -th component of the additive white Gaussian noise vector $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$, s is the transmitted information symbol selected from a constellation \mathcal{S} , such that $\mathbb{E}[s] = 0$ and $\mathbb{E}[|s|^2] = 1$. The real random variable, ϕ_m , is a phase noise increment, which is independent of s , θ_m , and the AWGN. Since $\phi_m \in [-\pi, \pi)$, a general representation is adopted for the probability density function (pdf), $p_{\Phi_m}(\phi_m)$, of ϕ_m , based on its Fourier expansion

$$p_{\Phi_m}(\phi_m) = \frac{1}{2\pi} \left(\alpha_{m,0} + 2 \sum_{l=1}^{\infty} \alpha_{m,l} \cos(l\phi_m) \right), \quad (3)$$

where $\alpha_{m,l}$, $l = 0, 1, 2, \dots$, are known real constants [27, 28]. Due to the wrapping of $\phi_m \in [-\pi, \pi)$ the pdf is periodic and the Fourier series expansion exists. The Fourier expansion in (3) can represent any pdf in $[-\pi, \pi)$ that is continuous, differentiable, unimodal, even and has zero mean. That includes the circular normal distribution (also known as von Mises or Tikhonov distribution) and the wrapped Gaussian distribution [29]. These two models are predominantly used in the literature to describe phase noise of practical oscillators. For free-running oscillators phase noise is often

modeled as a discrete-time Wiener process, where the increments are i.i.d. wrapped Gaussian increments [30]. For oscillators equipped with a PLL, the phase noise increment is well modeled by a random variable from a von Mises (or, equivalently, Tikhonov) distribution [11].

The setup presented in (1) and (2) can model two distinct operations, with respect to phase noise. In the first operation, the random variables θ_m and ϕ_m are independent across the BS antennas (*non-synchronous (NS)* operation). This models a practical distributed antenna deployment where the use of a separate oscillator per BS antenna is required. In the second operation, it holds that $\theta_1 \equiv \dots \equiv \theta_M$ and $\phi_1 \equiv \dots \equiv \phi_M$ (*synchronous (S)* operation). This corresponds to a practical centralized deployment where the same oscillator is used for the downconversion of the received passband signal to the baseband for all BS antennas. Hybrid topologies are also possible, however, they are less interesting to analyze since their SER is expected to lie between the SER of the S and NS operations. Hence, they are not considered in this work. Finally, (1) and (2) can be expressed in vector-matrix form as

$$\begin{aligned}\mathbf{x} &= \sqrt{\rho}\mathbf{\Theta}\mathbf{g} + \mathbf{w}, \\ \mathbf{y} &= \sqrt{\rho}\mathbf{\Theta}\mathbf{\Phi}\mathbf{g}s + \mathbf{z},\end{aligned}\quad (4)$$

where $\mathbf{x} \triangleq [x_1, \dots, x_M]^T$, $\mathbf{y} \triangleq [y_1, \dots, y_M]^T$, $\mathbf{\Theta} \triangleq \text{diag}\{e^{j\theta_1}, \dots, e^{j\theta_M}\}$, $\mathbf{\Phi} \triangleq \text{diag}\{e^{j\phi_1}, \dots, e^{j\phi_M}\}$, $\mathbf{g} \triangleq [g_1, \dots, g_M]^T$ is the vector of known amplitudes and $(\cdot)^T$ is the transpose operation.

2.2 Fading Channel (FC)

For the FC case, during the first channel use the signal received at the m -th BS antenna is given by

$$x_m = \sqrt{\rho}h_m + w_m, \quad (5)$$

where h_m is the m -th component of the zero mean circularly symmetric complex Gaussian channel vector, $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$. The initial phase θ_m that is present in (1) is absorbed in h_m . This can be assumed without modifying the statistics of h_m , due to the circular symmetry of the channel distribution. During the second channel use, the signal received at the m -th BS antenna is given by

$$y_m = \sqrt{\rho}e^{j\phi_m}h_ms + z_m. \quad (6)$$

The phase noise increment, ϕ_m , is defined as in (3). The fading coefficients h_m remain constant over both channel uses. Further, the S and NS operations are defined similarly as in Section 2.1. Finally, the vector–matrix form of (5) and (6) is given by

$$\begin{aligned}\mathbf{x} &= \sqrt{\rho}\mathbf{h} + \mathbf{w}, \\ \mathbf{y} &= \sqrt{\rho}\Phi\mathbf{h}s + \mathbf{z}.\end{aligned}\quad (7)$$

3 Optimal Detectors

In this section we describe the maximum likelihood (ML) detector for each of the four different cases described in Section 2. The *maximum a posteriori* (MAP) detector is identical to the ML detector, up to an additive constant dependent on the priors of s , and hence can be derived by trivial modification of the ML detectors. The optimal detectors are summarized in Table 1 for clarity. The BS uses the received vectors \mathbf{x} and \mathbf{y} jointly, to derive the optimum estimate, \hat{s} , of the transmitted information symbol, s , i.e.,

$$\hat{s} \triangleq \arg \max_{s \in \mathcal{S}} p(\mathbf{x}, \mathbf{y} | s). \quad (8)$$

We start with a proposition on the likelihood function for the NS operation.

Proposition 1. *The pdf of the received vectors (\mathbf{x}, \mathbf{y}) given a symbol s for the NS operation is given by*

$$p(\mathbf{x}, \mathbf{y} | s) = A \prod_{m=1}^M \left(\beta_{m,0} + 2 \sum_{l=1}^{\infty} \beta_{m,l} \cos(l\zeta_m) \right), \quad (9)$$

where for the CC-NS case we have

$$A = \exp \left(-\|\mathbf{x}\|^2 - \|\mathbf{y}\|^2 - \rho(1 + |s|^2) \|\mathbf{g}\|^2 \right) / \pi^{2M}, \quad (10)$$

$$\beta_{m,l} = \alpha_{m,l} I_l(2\sqrt{\rho}g_m |s^* y_m|) I_l(2\sqrt{\rho}g_m |x_m|), \quad (11)$$

$$\zeta_m = \arg(y_m) - \arg(x_m) - \arg(s) \quad (12)$$

and for the FC-NS case we have

$$A = \frac{e^{-\frac{1+\rho|s|^2}{\rho+\rho|s|^2+1}\|\mathbf{x}\|^2 - \frac{1+\rho}{\rho+\rho|s|^2+1}\|\mathbf{y}\|^2}}{(\pi^2(\rho + \rho|s|^2 + 1))^M} \quad (13)$$

$$\beta_{m,l} = \alpha_{m,l} I_l \left(\frac{2\rho |s^* x_m^* y_m|}{\rho + \rho|s|^2 + 1} \right) \quad (14)$$

and ζ_m as in (12). $I_l(\cdot)$ is the l -th order modified Bessel function of first kind [31], $(\cdot)^*$ is the complex conjugation operation and $\|\cdot\|$ is the Euclidean norm.

Proof. See Appendix 7.1 for the CC-NS case and Appendix 7.2 for the FC-NS case. \square

Based on the result in Proposition 1, the detector for the NS operation can be derived.

Corollary 1. For a discrete constellation, \mathcal{S} , the ML symbol, \hat{s} , for the NS operation

$$\begin{aligned} \hat{s} &= \arg \max_{s \in \mathcal{S}} \mathcal{L}_s^{NS} \\ &= \arg \max_{s \in \mathcal{S}} B + \sum_{m=1}^M \ln \left(\beta_{m,0} + 2 \sum_{l=1}^{\infty} \beta_{m,l} \cos(l\zeta_m) \right). \end{aligned} \quad (15)$$

where \mathcal{L}_s^{NS} is the decision metric for the symbol $s \in \mathcal{S}$. For the CC-NS case we have

$$B = -\rho|s|^2 \|\mathbf{g}\|^2, \quad (16)$$

$\beta_{m,l}$ as in (11) and ζ_m as in (12). For the FC-NS case we have

$$B = -M \ln(1 + \rho + \rho|s|^2) - \frac{1 + \rho|s|^2}{1 + \rho + \rho|s|^2} \|\mathbf{x}\|^2 - \frac{1 + \rho}{1 + \rho + \rho|s|^2} \|\mathbf{y}\|^2, \quad (17)$$

$\beta_{m,l}$ as in (14) and ζ_m as in (12).

Proof. The result follows trivially from Proposition 1 by taking the natural logarithm of (44), $\ln(p(\mathbf{x}, \mathbf{y}|s))$, and dropping the terms that are independent of s . \square

The results in Corollary 1 hold for arbitrary constellations. In the following we particularize Corollary 1 for the case of phase shift keying (N -PSK) constellations. A rigorous motivation for this choice is deferred to Section 4.1.

Corollary 2. For s selected from an N -PSK constellation (i.e., $s \in \{e^{j\frac{2\pi n}{N}}\}_{n=0}^{N-1}$) for the NS operation, the ML detection rule is given by (15) with B as in (16) for $|s|^2 = 1$

$$\begin{aligned}\beta_{m,l} &= \alpha_{m,l} I_l(2\sqrt{\rho}g_m|x_m|) I_l(2\sqrt{\rho}g_m|y_m|) \text{ and} \\ \zeta_m &= \arg(y_m) - \arg(x_m) - \frac{2\pi n}{N}\end{aligned}\quad (18)$$

for the CC-NS case and B as in (17) for $|s|^2 = 1$

$$\beta_{m,l} = \alpha_{m,l} I_l\left(\frac{2\rho|x_m^*y_m|}{2\rho+1}\right) \text{ and } \zeta_m = \arg(y_m) - \arg(x_m) - \frac{2\pi n}{N} \quad (19)$$

for the FC-NS case.

The counterparts of Proposition 1 and Corollaries 1 and 2 for the S operation are provided in the following.

Proposition 2. The pdf of the received vectors (\mathbf{x}, \mathbf{y}) given a symbol s for the S operation is given by

$$p(\mathbf{x}, \mathbf{y}|s) = A \left(\beta_0 + 2 \sum_{l=1}^{\infty} \beta_l \cos(l\zeta) \right) \quad (20)$$

where for the CC-S case we have A as in (10),

$$\beta_l = \alpha_l I_l(2\sqrt{\rho}|s^* \mathbf{g}^T \mathbf{y}|) I_l(2\sqrt{\rho}|\mathbf{g}^T \mathbf{x}|), \quad (21)$$

$$\zeta = \arg(\mathbf{g}^T \mathbf{y}) - \arg(\mathbf{g}^T \mathbf{x}) - \arg(s) \quad (22)$$

and for the FC-S case we have A as in (13),

$$\beta_l = \alpha_l I_l\left(\frac{2\rho|s^* \mathbf{x}^H \mathbf{y}|}{1 + \rho + \rho|s|^2}\right), \quad (23)$$

$$\zeta = \arg(\mathbf{x}^H \mathbf{y}) - \arg(s) \quad (24)$$

where $(\cdot)^H$ is the complex conjugation and transposition operation.

Proof. See Appendix 7.3 for the CC-S case and Appendix 7.4 for the FC-S case. \square

The detector for the S operation is given by Corollary 3.

	CC	FC
NS	Corollary 1	Corollary 1
	(15), (16), (11), (12)	(15), (17), (14), (12)
S	Corollary 3	Corollary 3
	(25), (16), (21), (22)	(25), (17), (23), (24)

Table 1: Summary of optimal detection rules for all the cases under investigation.

Corollary 3. For a discrete constellation, \mathcal{S} , the ML symbol, \hat{s} , for the S operation is

$$\hat{s} = \arg \max_{s \in \mathcal{S}} \mathcal{L}_s^S = \arg \max_{s \in \mathcal{S}} B + \ln \left(\beta_0 + 2 \sum_{l=1}^{\infty} \beta_l \cos(l\zeta) \right). \quad (25)$$

where \mathcal{L}_s^S is the decision metric for the symbol $s \in \mathcal{S}$. For the CC-S case we have B as in (16), β_l as in (21) and ζ as in (22). For the FC-S case we have B as in (17), β_l as in (23) and ζ as in (24).

Proof. The proof is done similarly to the proof of Corollary 1. \square

For an N -PSK constellation the detector in Corollary 3 can be particularized as in Corollary 4.

Corollary 4. For s selected from an N -PSK constellation (i.e. $s \in \{e^{j\frac{2\pi n}{N}}\}_{n=0}^{N-1}$) for the synchronous operation, the ML detection rule is given by (25) with B as in (16) for $|s|^2 = 1$

$$\begin{aligned} \beta_l &= \alpha_l I_l(2\sqrt{\rho}|\mathbf{g}^T \mathbf{x}|) I_l(2\sqrt{\rho}|\mathbf{g}^T \mathbf{y}|) \text{ and} \\ \zeta &= \arg(\mathbf{g}^T \mathbf{y}) - \arg(\mathbf{g}^T \mathbf{x}) - \frac{2\pi n}{N} \end{aligned} \quad (26)$$

for the CC-S case and B as in (17) for $|s|^2 = 1$

$$\beta_l = \alpha_l I_l \left(\frac{2\rho|\mathbf{x}^H \mathbf{y}|}{2\rho + 1} \right) \text{ and } \zeta = \arg(\mathbf{x}^H \mathbf{y}) - \frac{2\pi n}{N} \quad (27)$$

for the FC-S case.

3.1 Implementation of (15) and (25)

At this point, a comment on computational complexity and implementation issues of the detectors in Corollaries 1 and 3 is in order. For the S operation, the calculation of the optimal detectors requires the computation of inner products between vectors of size M . For the NS operation, the computational complexity scales also linearly with M . Hence, the complexity of the optimal detectors scales in all cases linearly with the number of BS antennas. From (15) and (25), the ML detectors involve the calculation of infinite series, where the l -th term is a function of the modified Bessel functions, $I_l(\cdot)$. However, the implementation of these detectors is still feasible by appropriate truncation of the infinite series. Since the series converge very fast, the truncation error can be made negligible. The following lemma is useful to demonstrate this claim.

Lemma 1 ([32,33]). *For $\mu > \nu > 0$ and $x > 0$ it holds that*

$$\frac{I_\nu(x)}{I_\mu(x)} > \max \left\{ 1, \left(\frac{x}{2} \right)^{\nu-\mu} \frac{\Gamma(\mu + \frac{1}{2})}{\Gamma(\nu + \frac{1}{2})} \right\}, \quad (28)$$

where $\Gamma(\cdot)$ is the Gamma function $\Gamma(z) \triangleq \int_0^\infty e^{-t} t^{z-1} dt$, defined for $\Re\{z\} > 0$ [34, 8.310].

The ratio $\frac{I_\mu(x)}{I_1(x)}$ and the bound in Lemma 1 are plotted as a function of the argument x for various values of μ in Fig. 2. For all $\mu > 1$ and x it holds that $\frac{I_\mu(x)}{I_1(x)} < 1$ and for small to moderate values of x we even have $\frac{I_\mu(x)}{I_1(x)} \ll 1$. Further, the ratio $\frac{I_\mu(x)}{I_1(x)}$ is monotonically decreasing in μ when x is fixed. In fact the rate of decrease is quite fast, since, for any integer μ , $\Gamma(\mu) = (\mu - 1)!$ and x^μ grows at a lesser rate than the factorial function. The rate of decrease as a function of μ for fixed x is shown in Fig. 3. This fast convergence establishes the fact that only a few terms of the infinite sums are required so that the approximation error is negligible.

3.2 The ML Detector for von Mises Phase Noise Increments

We conclude the section by particularizing the results in a special, but important, case, where the optimal detector structure does not require the implementation of an infinite series. This involves the FC-NS and FC-S cases.

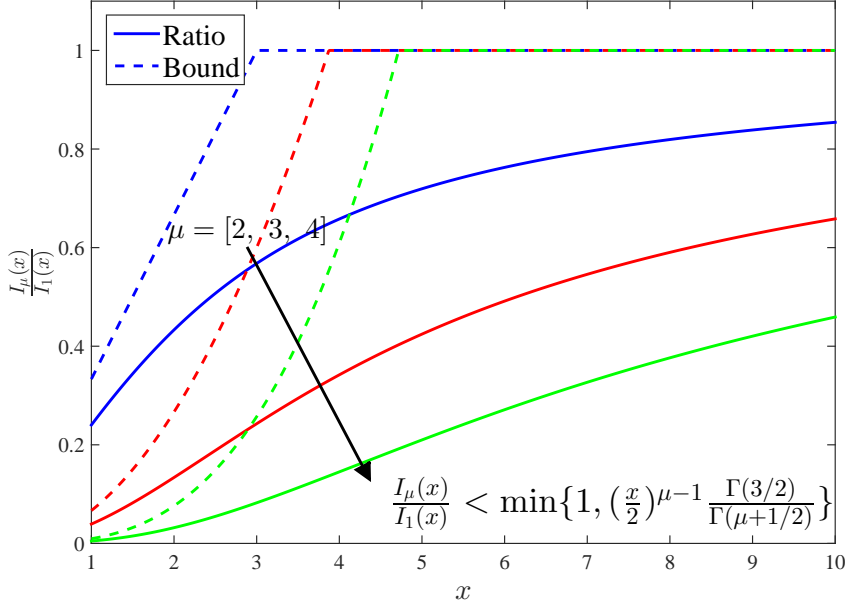


Figure 2: Ratio of modified Bessel functions $\frac{I_\mu(x)}{I_1(x)}$ and the bound from Lemma 1 for $\mu = [2, 3, 4]$ as a function of x .

Proposition 3. For the fading case with $\phi_m \sim \mathcal{VM}(0, \kappa)$, distributed as zero mean von Mises random variables with concentration parameter κ , the likelihood functions can be expressed for the NS operation as

$$p(\mathbf{x}, \mathbf{y}|s) = A \prod_{m=1}^M \frac{I_0 \left(\sqrt{\kappa^2 + (\mathbf{b}_m^{NS})^2 + 2\kappa (\mathbf{b}_m^{NS}) \cos(\zeta_m)} \right)}{I_0(\kappa)} \quad (29)$$

and for the S operation as

$$p(\mathbf{x}, \mathbf{y}|s) = A \frac{I_0 \left(\sqrt{\kappa^2 + (\mathbf{b}^S)^2 + 2\kappa (\mathbf{b}^S) \cos(\zeta)} \right)}{I_0(\kappa)}, \quad (30)$$

where A is defined as in (13), $\mathbf{b}_m^{NS} = \frac{2\rho|s^*x_m^*y_m|}{\rho+\rho|s|^2+1}$, $\mathbf{b}^S = \frac{2\rho|s^*\mathbf{x}^H\mathbf{y}|}{\rho+\rho|s|^2+1}$, ζ_m as in (12) and ζ as in (24).

Proof. See Appendix 7.5. □

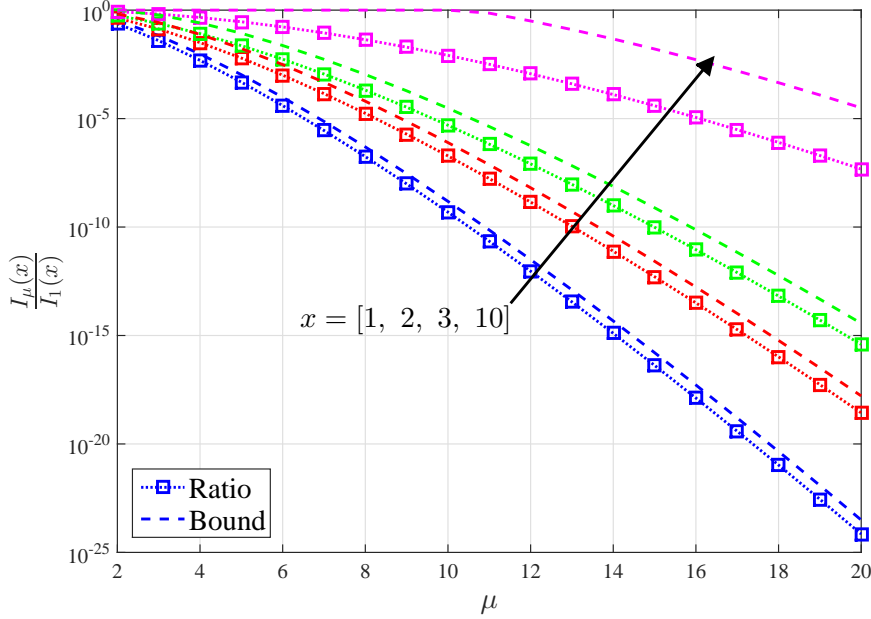


Figure 3: Ratio of modified Bessel functions $\frac{I_\mu(x)}{I_1(x)}$ and the bound from Lemma 1 for a set of fixed values of x as a function of μ .

4 High SNR Analysis

4.1 High SNR Analysis for the Synchronous Operation

The expressions in Propositions 1, 2 and Corollaries 1-4 can be easily implemented, however, it is hard to extract insight on the fundamental behavior of the different operations. Therefore, in this section we present an asymptotic analysis as $\rho \rightarrow \infty$ (high-SNR) for the system models in (4) and (7) in order to reveal the high-SNR behavior of the detectors. We start with the synchronous operation as it appears to be simpler. The system model for the CC-S as $\rho \rightarrow \infty$ can be expressed as

$$\tilde{\mathbf{y}} \triangleq \lim_{\rho \rightarrow \infty} \frac{\mathbf{y}}{\sqrt{\rho}} = e^{j(\theta+\phi)} s \mathbf{g} \Rightarrow \tilde{\mathbf{y}} = \tilde{\mathbf{x}} e^{j\phi} s \Rightarrow \begin{cases} |\tilde{\mathbf{x}}^H \tilde{\mathbf{y}}| = \|\mathbf{g}\|^2 |s| \\ \psi = \arg(s) + \phi \end{cases} \quad (31)$$

where $\tilde{\mathbf{x}} \triangleq \lim_{\rho \rightarrow \infty} \frac{\mathbf{x}}{\sqrt{\rho}} = e^{j\theta} \mathbf{g}$ and $\psi \triangleq \arg(\tilde{\mathbf{x}}^H \tilde{\mathbf{y}})$. From (31) it is apparent that the amplitude of s can be decoded error-free. Hence, distinguishing

two symbols that have different amplitude is trivial at high-SNR. However, distinguishing two symbols that have different phase is much more challenging. Therefore, in the following, we restrict the study to constellations that encode information only in the phase, that is PSK. Two-dimensional constellations, such as 16-QAM or 64-QAM, can be treated as consisting of multiple disjoint sub-constellations, where the information is conveyed only on the phase as explained in the following. Points of the two dimensional constellation that are at different distances from the origin belong to different sub-constellations with different radii and hence are easily distinguishable at high-SNR. Constellation points that are at the same radius belong to the same sub-constellation and their pairwise error probability can be calculated in a manner similar to PSK constellations. The observation ψ in (31) is in this case sufficient statistics for the detection of $\arg(s)$. We proceed by deriving the asymptotic SER at high-SNR for PSK constellations.

Proposition 4. *For the CC-S case, let ϕ be zero mean random variable with pdf $p_{\Phi}(\phi)$ as in (3), where the distribution is unimodal and symmetric around the mean.¹ Define the error event $\epsilon = \{\arg(\hat{s}) \neq 0 | \arg(s) = 0\}$. Then the SER floor at the high-SNR for equiprobable N -PSK symbols is given by*

$$\Pr\{\epsilon\} \triangleq 1 - \int_{-\frac{\pi}{N}}^{\frac{\pi}{N}} p_{\Phi}(\phi) d\phi = 1 - \frac{\alpha_0}{N} - \sum_{l=1}^{\infty} \frac{2\alpha_l}{\pi l} \sin\left(l \frac{\pi}{N}\right). \quad (32)$$

Corollary 5. *From (32) we observe that there is a non-zero SER floor for the CC-S case, which depends only on the statistics of the phase noise increment and the PSK constellation density, N , but is independent of the number of receive antennas, M .*

Remark 1. *The preceding analysis is also true for FC-S by defining $\tilde{\mathbf{x}} \triangleq \mathbf{h}$ and $\tilde{\mathbf{y}} \triangleq \mathbf{h}e^{j\phi}s$. Observe that the statistic corresponding to the amplitude (31) is now $|\tilde{\mathbf{x}}^H \tilde{\mathbf{y}}| = \|\mathbf{h}\|^2 |s|$, where \mathbf{h} is stochastic. This generally has an impact of the performance in FC-S, however, in the high-SNR regime the effect of the phase noise increment dominates and the analysis for CC-S applies in FC-S, as well. The performance of FC-S approaches that of CC-S as $M \rightarrow \infty$ due to the hardening of $\frac{1}{M} \|\mathbf{h}\|^2 \xrightarrow{M \rightarrow \infty} \mathbb{E}[|h_1|^2]$ almost surely [35].*

¹This assumption is not restrictive in practice as the widely accepted models on the phase noise impairments, such as the Wiener model, satisfy it.

4.2 High SNR Analysis for the CC-NS Case

For the CC-NS case based on the model in (1) and (2) we define $\tilde{x}_m \triangleq \lim_{\rho \rightarrow \infty} \frac{x_m}{\sqrt{\rho}} = e^{j\theta_m}$ and $\tilde{y}_m \triangleq \lim_{\rho \rightarrow \infty} \frac{y_m}{\sqrt{\rho}}$. By proceeding similarly to Section 4.1 it can be shown that the amplitude can be decoded error free so we concentrate on the transmitted phase, which for the m -th BS antenna element is given by $\psi_m \triangleq \arg(\tilde{x}_m^* \tilde{y}_m) = \phi_m + \arg(s)$. The general parameterization of the phase noise increment in (3) does not yield mathematically tractable expressions. To get a mathematically tractable expression, we consider that the increments ϕ_m are independent $\mathcal{VM}(0, \kappa)$ random variables for $m = 1, \dots, M$. As noted earlier, the von Mises distribution is used to describe phase noise resulting from oscillators that are equipped with PLL. It is symmetric around its mean, unimodal and is representative of the anticipated statistical behavior of phase noise. Hence, we expect that the general insights drawn from this choice will hold for other similar distributions. The likelihood function of the observed phases $\boldsymbol{\psi} \triangleq [\psi_1, \dots, \psi_M]^T$, given the transmitted phase $\arg(s)$, can be expressed as

$$p_{\boldsymbol{\Psi}|S}(\boldsymbol{\psi}|s) = \prod_{m=1}^M p_{\Psi_m|s}(\psi_m|s) = \frac{e^{\kappa \sum_{m=1}^M \cos(\psi_m - \arg(s))}}{(2\pi I_0(\kappa))^M}$$

and the corresponding ML decision rule for symbols selected from some alphabet \mathcal{S} is given by

$$\hat{s} = \arg \max_{s \in \mathcal{S}} p_{\boldsymbol{\Psi}|S}(\boldsymbol{\psi}|s) = \arg \max_{s \in \mathcal{S}} \sum_{m=1}^M \cos(\psi_m - \arg(s)). \quad (33)$$

Proposition 5. *The decision metric, μ_n , for the symbol $e^{j\frac{2\pi n}{N}}$ from an N -PSK constellation based on (33) is given by*

$$\begin{aligned} \mu_n &\triangleq \sum_{m=1}^M \left(\cos\left(\psi_m - \frac{2\pi n}{N}\right) - \cos(\psi_m) \right) \\ &= \sum_{m=1}^M \sin\left(\frac{\pi n}{N}\right) \sin\left(\psi_m - \frac{\pi n}{N}\right). \end{aligned} \quad (34)$$

If we denote by ϵ the error event, i.e., the case where the detected symbol \hat{s} is different from the transmitted symbol s , then the SER for equiprobable input symbols is

given by

$$\Pr \{ \epsilon \} = \Pr \left\{ \bigcup_{n=1}^{N-1} \{ \mu_n > 0 \} \mid \arg(s) = 0 \right\}. \quad (35)$$

The exact calculation of the probability of error in (35) appears formidable. We therefore derive an upper bound on the pairwise symbol error probability of erroneously detecting $s_n = \exp(j\frac{2\pi n}{N})$, $n = 1, \dots, N-1$ when $s_0 = 1$ was sent. We note that due to the symmetry of the von Mises distribution around its mean and the uniform priors on the input symbols, the conditioning on any particular input symbol does not affect the result. The symbol $s_0 = 1$ with $\arg(s_0) = 0$ is selected for convenience. From (34) the μ_n is a sum of bounded independent and identically distributed (i.i.d.) random variables $Z_m \triangleq \sin(\frac{\pi n}{N}) \sin(\psi_m - \frac{\pi n}{N})$. Hence the following lemma can be used.

Lemma 2 (Bernstein Inequality [36]). *Let X_m , $m = 1, \dots, M$, be i.i.d. random variables with $\mathbb{E}[X_m] = 0$, $|X_m| < C$ almost surely for some bounded constant C , $X_s \triangleq \sum_{m=1}^M X_m$ and $\varsigma \triangleq \sqrt{\text{VAR}(X_s)}$. Then for all $t > 0$*

$$\Pr \{ X_s > t\varsigma \} \leq \exp \left(-\frac{t^2}{2 + \frac{2}{3} \frac{C}{\varsigma} t} \right).$$

Proposition 6. *The pairwise error probability for the detected symbol \hat{s}_n to be $s_n = \exp(j\frac{2\pi n}{N})$, $n = 1, \dots, N-1$, given that the symbol $s_0 = 1$ was sent is upper bounded by*

$$\Pr \{ \mu_n > 0 \mid s = 1 \} \leq \exp \left(-\frac{M \left(\frac{\sin^2(\frac{\pi n}{N}) I_1(\kappa)}{\sqrt{\text{VAR}(X_{m,n})} I_0(\kappa)} \right)^2}{2 + \frac{2}{3} \frac{C \sin^2(\frac{\pi n}{N}) I_1(\kappa)}{\text{VAR}(X_{m,n}) I_0(\kappa)}} \right), \quad (36)$$

where $C \triangleq \sin(\frac{\pi n}{N}) + \sin^2(\frac{\pi n}{N}) \frac{I_1(\kappa)}{I_0(\kappa)}$ and

$$\text{VAR}(X_{m,n}) = \sin^2\left(\frac{\pi n}{N}\right) \left(\frac{I_1(\kappa) \cos\left(\frac{2\pi n}{N}\right)}{\kappa I_0(\kappa)} + \sin^2\left(\frac{\pi n}{N}\right) \left(1 - \frac{I_1^2(\kappa)}{I_0^2(\kappa)} \right) \right). \quad (37)$$

Proof. Let $X_{m,n} \triangleq \sin\left(\frac{\pi n}{N}\right) \sin\left(\psi_m - \frac{\pi n}{N}\right) + \sin^2\left(\frac{\pi n}{N}\right) \frac{I_1(\kappa)}{I_0(\kappa)}$. Then $\mathbb{E}[X_{m,n}] = 0$, $|X_{m,n}| \leq C \triangleq \sin\left(\frac{\pi n}{N}\right) + \sin^2\left(\frac{\pi n}{N}\right) \frac{I_1(\kappa)}{I_0(\kappa)}$ and $\text{VAR}(X_{m,n})$ is given by (37). Then

$$\begin{aligned} \Pr\{\mu_n > 0 | s = 1\} &= \Pr\left\{\sum_{m=1}^M \sin\left(\frac{\pi n}{N}\right) \sin\left(\psi_m - \frac{\pi n}{N}\right) > 0\right\} \\ &= \Pr\left\{\sum_{m=1}^M X_{m,n} > M \sin^2\left(\frac{\pi n}{N}\right) \frac{I_1(\kappa)}{I_0(\kappa)}\right\}. \end{aligned}$$

Define $\varsigma \triangleq \sqrt{M} \sqrt{\text{VAR}(X_{m,n})}$ and $t \triangleq \sqrt{M} \frac{\sin^2\left(\frac{\pi n}{N}\right) \frac{I_1(\kappa)}{I_0(\kappa)}}{\sqrt{\text{VAR}(X_{m,n})}}$. The result in (36) follows by application of Lemma 2. \square

Proposition 6 shows that the pairwise error probability for the CC-NS can be reduced exponentially with M . In addition, due to the union bound, the SER floor in (35) can be reduced arbitrarily to zero by increasing M .

4.3 High SNR Analysis for the FC-NS Case

Similarly to Remark 1, for the FC-NS case the following variables are defined $\tilde{x}_m \triangleq \lim_{\rho \rightarrow \infty} \frac{x_m}{\sqrt{\rho}}$ and $\tilde{y}_m \triangleq \lim_{\rho \rightarrow \infty} \frac{y_m}{\sqrt{\rho}}$, which in the high-SNR regime simplify to $\tilde{x}_m = h_m$ and $\tilde{y}_m = e^{j\phi_m} h_m s$. The observation vector $\mathbf{v} = [v_1, \dots, v_m, \dots, v_M]^T$ with $v_m \triangleq \tilde{x}_m^* \tilde{y}_m = |h_m|^2 e^{j\phi_m} s$ is further defined. The derivation of the optimal detection rule even in this regime appears to be mathematically intractable. We seek to find an upper bound on the high-SNR SER floor by using a heuristic suboptimal decision statistic. We assume that ϕ_m are i.i.d. $\mathcal{VM}(0, \kappa)$ for some $\kappa > 0$ random variables and we define the decision vector $\zeta \triangleq [\zeta_1 \ \zeta_2]^T$, where

$$\begin{aligned} \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix} &\triangleq \frac{1}{M} \begin{bmatrix} \Re \left\{ \sum_{m=1}^M v_m \right\} \\ \Im \left\{ \sum_{m=1}^M v_m \right\} \end{bmatrix} \\ &= \frac{1}{M} |s|^2 \begin{bmatrix} \sum_{m=1}^M |h_m|^2 \cos(\phi_m + \arg(s)) \\ \sum_{m=1}^M |h_m|^2 \sin(\phi_m + \arg(s)) \end{bmatrix}. \end{aligned} \quad (38)$$

The suboptimal decision rule that we use is the minimum Euclidean distance from a scaled N -PSK, i.e.,

$$\hat{s} = \arg \min_{s \in \mathcal{S}} \|\zeta - s\|, \quad (39)$$

where \mathcal{S} is the N -PSK alphabet. Conditioned on the symbol $s_0 = [1 \ 0]^T$ being sent and given the decision rule (39), an error occurs when $\epsilon = \{\hat{s} \neq s_0 | s = s_0\}$. Hence, the symbol error probability can be upper bounded by the union bound as follows:

$$\Pr \{\epsilon\} \leq \sum_{n=1}^{N-1} \Pr \{\hat{s} = s_n | s = s_0\} \quad (40)$$

where $s_n = [\cos(\frac{2\pi n}{N}) \ \sin(\frac{2\pi n}{N})]^T$. We provide here a lemma that is useful to bound (40).

Lemma 3 (Chebychev's Inequality [37]). *Let X be a random variable with mean μ and variance σ^2 . For any $\varepsilon > 0$,*

$$\Pr \{|X - \mu| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}. \quad (41)$$

Corollary 6. *The SER floor at high-SNR for the FC-NS case in (40) scales at least as $O(\frac{1}{M})$.*

Proof. The pairwise error probabilities, $\Pr \{\hat{s} = s_n | s = s_0\}$, are bounded as follows

$$\begin{aligned} \Pr \{\hat{s} = s_n | s = s_0\} &= \Pr \{\|s_0 - \zeta\| > \|s_n - \zeta\| | s = s_0\} \\ &= \Pr \{\zeta^T (s_0 - s_n) < 0 | s = s_0\} = \Pr \{\xi_n < 0\} \\ &= \Pr \{-\xi_n + \mathbb{E}[\xi_n] > \mathbb{E}[\xi_n]\} \stackrel{(a)}{\leq} \Pr \{|\xi_n - \mathbb{E}[\xi_n]| > \mathbb{E}[\xi_n]\} \stackrel{(b)}{\leq} \frac{\text{VAR}(\xi_n)}{(\mathbb{E}[\xi_n])^2} \end{aligned}$$

where

$$\xi_n \triangleq \frac{1 - \cos(\frac{2\pi n}{N})}{M} \sum_{m=1}^M |h_m|^2 \cos(\phi_m) + \frac{\sin(\frac{2\pi n}{N})}{M} \sum_{m=1}^M |h_m|^2 \sin(\phi_m)$$

and $\mathbb{E}[\xi_n] = (1 - \cos(\frac{2\pi n}{N})) \frac{I_1(\kappa)}{I_0(\kappa)}$. The inequality in (a) follows from the fact that $\{-\xi_n + \mathbb{E}[\xi_n] > \mathbb{E}[\xi_n]\} \subseteq \{|\xi_n - \mathbb{E}[\xi_n]| > \mathbb{E}[\xi_n]\}$ and (b) follows from

Lemma 3. Calculation of the variance $\text{VAR}(\xi_n)$ gives

$$\begin{aligned} \text{VAR}(\xi_n) &= \frac{(1 - \cos(\frac{2\pi n}{N}))^2}{M} \left(1 + \frac{I_2(\kappa)}{I_0(\kappa)} - \left(\frac{I_1(\kappa)}{I_0(\kappa)} \right)^2 \right) \\ &\quad + \frac{(\sin(\frac{2\pi n}{N}))^2}{M} \left(1 - \frac{I_2(\kappa)}{I_0(\kappa)} \right). \end{aligned}$$

Hence a positive constant c independent of M can be found such that $\text{VAR}(\xi_n) \leq c \frac{1}{M} \Rightarrow \text{VAR}(\xi_n) = O\left(\frac{1}{M}\right)$. \square

Since the pairwise error probabilities are $O\left(\frac{1}{M}\right)$, from (40) the SER floor at the high-SNR for the FC-NS case is also $O\left(\frac{1}{M}\right)$. Hence, the SER floor for the FC-NS case can be made arbitrarily small as $M \rightarrow \infty$.

We conclude this section with a short intuitive explanation for the SER floors in both operations. At high-SNR in the S operation every BS antenna observes exactly the same signal. As a result, no advantage can be gained by using multiple antennas. In contrast, in the NS operation each BS antenna observes the symbol perturbed by some independent phase rotation. Hence, an averaging effect is observed for the independent phase noise sources at the BS array. Similar conclusions were also drawn in [11] for the uplink case.

We note that the single antenna transmitter is assumed to be phase noise free as our concern is the effect of the phase noise impairments at the BS array. The results are still valid in the case of a phase noise impaired transmitter for the S operation with appropriate adjustment of the notation. However, in the NS operation the averaging effect will not appear for the phase noise at the transmitter. Similar conclusions have already been drawn in prior work, such as [12], where it is shown that in the NS operation the phase noise impairments at the BS average out but not the ones at the user terminals.

5 Numerical Examples

In this section we present numerical examples that verify the validity of the analytical results presented in Sections 3 and 4. The examples are obtained by Monte Carlo simulations, where the receiver uses the detectors

from Section 3 to decide on the transmitted information symbols. In Fig. 4 the zero mean von Mises distribution is plotted for reference purposes for various values of the concentration parameter κ . It is observed that the distribution is unimodal and that for $\kappa = 0$ it corresponds to the uniform distribution. As κ increases the distribution becomes more concentrated around the mean. In the following we will refer to this distribution for our results, even though all the propositions and corollaries in Section 3 hold for distributions that can be parameterized as in (3).

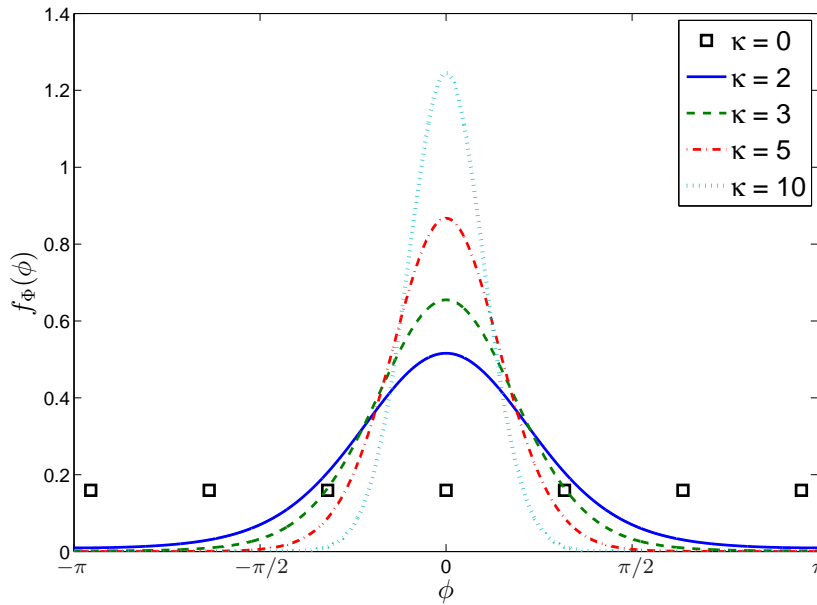


Figure 4: The von Mises distribution, $\mathcal{VM}(0, \kappa)$, for various choices of the parameter κ .

In Fig. 5 the SER performance as a function of the SNR ρ [dB] is plotted for $\kappa = 4$ and $M \in \{2, 4, 6\}$ for the constant channel case. In all the numerical examples for the CC case is assumed to be $\mathbf{g} = [1, \dots, 1]^T$. In the low SNR regime array gains are observed for both operations and the synchronous operation is marginally better than the non-synchronous operation. However, the performance of both operations in that regime prohibits reliable communication. In the medium SNR regime ($\approx 0 - 10$ [dB]) the non-synchronous operation has a clear advantage over the synchronous operation. SER floors in the high-SNR regime are observed in all cases.

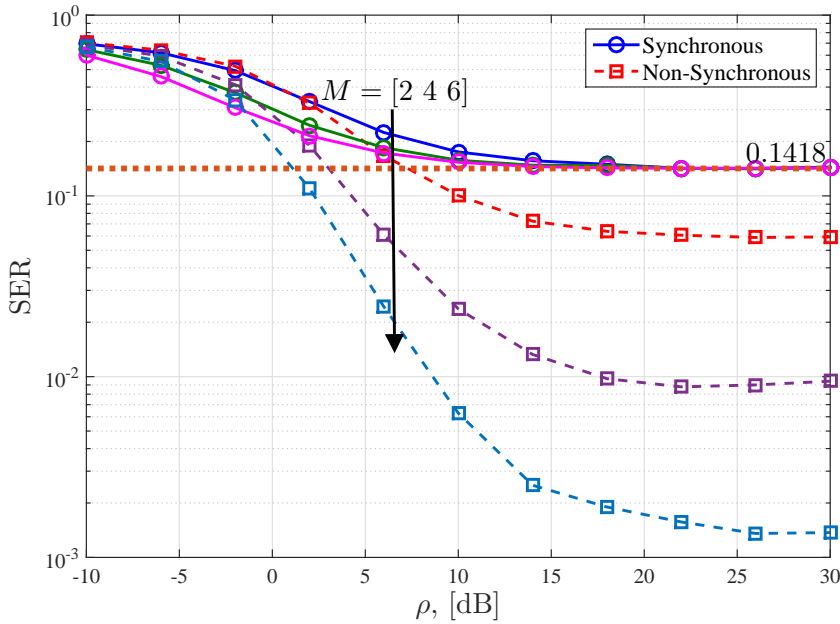


Figure 5: SER performance as a function of the SNR ρ for the constant channel case. The symbols are selected from a QPSK constellation and the concentration parameter is $\kappa = 4$ for various values of M .

Specifically for the synchronous operation the SER performance saturates at the same value irrespective of M . This is in line with Proposition 4. The dotted straight line corresponds to the theoretically calculated SER floor, (32). The theoretical value 0.1418 is also shown. Superior high-SNR performance is observed for the non-synchronous operation. In this case the SER floor is reduced when increasing the number of BS antennas. This verifies the results in Sections 4.2 and 4.3.

In Fig. 6 the SER performance as a function of ρ [dB] is plotted for $\kappa = 4$ and $M \in \{2, 4, 6\}$ for the fading channel case. For the synchronous operation, an array gain is observed in the low SNR regime and the SER floor is the same for all the values of M . These observations are identical to the constant channel case and are in line with Proposition 4. The theoretical SER floor is also plotted as in Fig. 5. Analogously to the constant channel case, SER floors are observed in the FC-NS case as well. Further, the SER floor is reduced by increasing the number of BS antennas. This is also in line with the results in Sections 4.2 and 4.3.

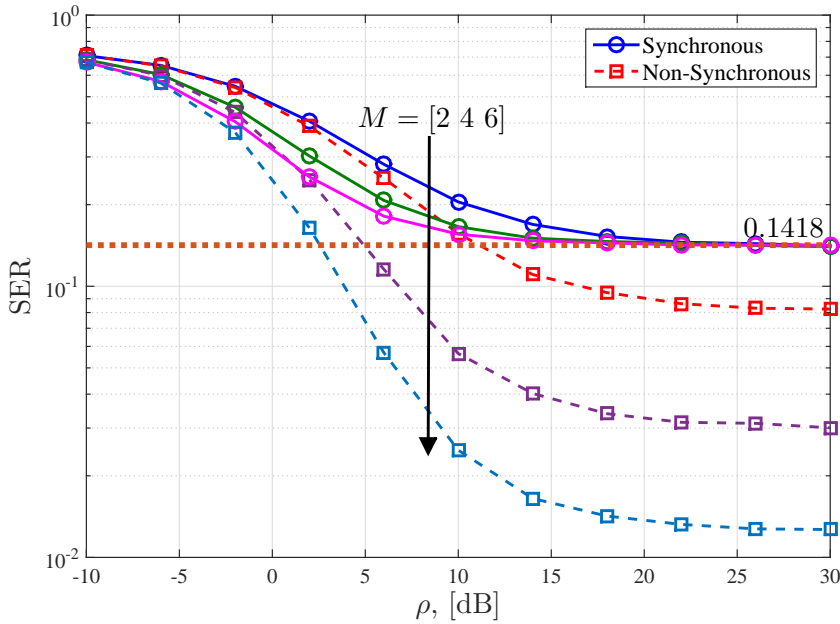


Figure 6: SER performance as a function of the SNR ρ for the fading channel case. The symbols are selected from a QPSK constellation and the concentration parameter is $\kappa = 4$ for various values of M .

For the Figs. 5 and 6 the detectors in (15) and (25) were computed so that the relative accuracy, $\delta_{\text{acc}}(\nu)$, of the truncated metric with ν terms for the s constellation symbol, $\tilde{\mathcal{L}}_s(\nu)$, defined as $\delta_{\text{acc}}(\nu) \triangleq \left| \frac{\tilde{\mathcal{L}}_s(\nu) - \tilde{\mathcal{L}}_s(\nu-1)}{\tilde{\mathcal{L}}_s(\nu-1)} \right|$, was less than 10^{-12} . In Table 2 we summarize the mean and the maximum number of terms required to achieve the relative accuracy, $\delta_{\text{acc}}(\nu)$, for a set of values of ρ [dB]. We observe that as ρ increases the number of terms required increases as well. However, in all cases less than 20 terms were sufficient for the desired relative accuracy. This demonstrates the fact that with only a few terms included in (15) and (25), the truncation error becomes negligible.

In Fig. 7 the SER performance is plotted as a function of the SNR ρ [dB] for the constant channel and fading channel cases. Similar performance is observed for the synchronous operation in both CC-S and FC-S cases. In the low SNR the constant channel exhibits better performance. This is due to the randomness of the fading channel case. However, as the number of

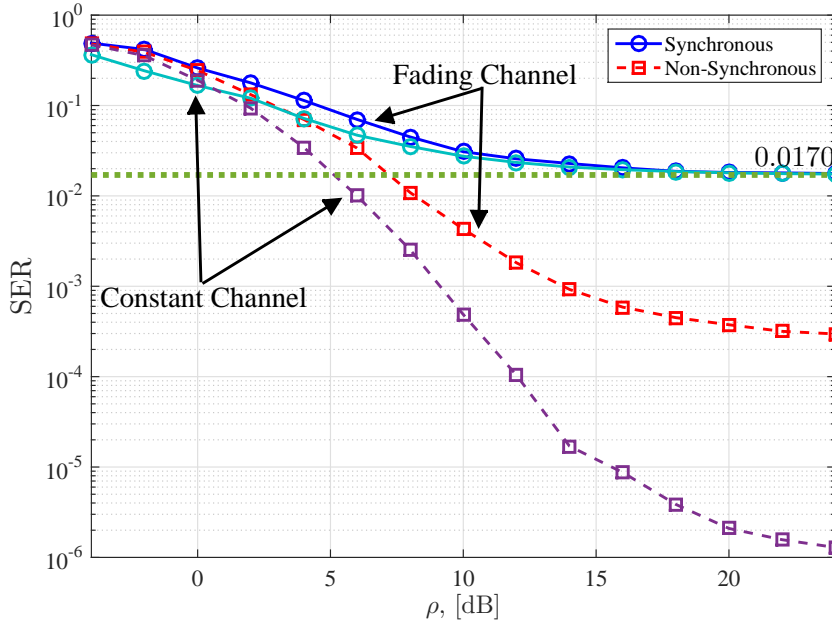


Figure 7: Comparison of the SER performance as a function of the SNR ρ for the CC and FC cases. The symbols are selected from a QPSK constellation and the concentration parameter is $\kappa = 10$ and for $M = 5$ BS antennas.

BS antennas increases the squared norm of the fading channel vector normalized by M becomes almost deterministic, so this gap is expected to be reduced. In the high-SNR regime the SER floor is the same for both the constant channel and the fading channel case. This is in agreement with Remark 1. The theoretical SER floor for the CC-S and FC-S cases is also given by the dotted line. For the non-synchronous operation improved performance is observed in the constant channel case. This implies that the additional randomness due to fading in the FC-NS case has a direct impact on the SER performance. This is an expected behavior. In the absence of phase noise the CC-NS case corresponds to the detection in an AWGN channel with constant gain and the error probability can be expressed in terms of the Q -function. In the absence of phase noise, the FC-NS case corresponds to the coherent detection in a Rayleigh SIMO channel with perfect channel knowledge and the error probability scales only as ρ^{-M} [38]. This phenomenon naturally carries over when phase noise impairments are present.

		2 dB		10 dB		22 dB	
		CC	FC	CC	FC	CC	FC
Mean	S	13.8	12.7	16.2	16.2	15.7	17.1
	NS	10	12.5	13.8	16	15.8	16.6
Max	S	17	16	17	17	16	18
	NS	13	16	17	17	17	18

Table 2: Mean and maximum number of terms, ν , in the truncated sums of the optimal detectors from (15) and (25) for relative accuracy of $\delta_{\text{acc}}(\nu) < 10^{-12}$, $M = 6$ and $\kappa = 4$.

5.1 Extension to Longer Data Intervals

In practice more than one channel uses are spent for data transmission. In this section we demonstrate that the conclusions drawn by the study of the models in (4) and (7) are valid for transmission protocols with multiple channel uses. For this purpose, we consider a setup where the data interval is extended to T channel uses. The calculation of the optimal detectors in closed form appears to be intractable [16], [18], [19], [21]. Hence, the approach can be summarized as follows. If a suboptimal tractable detector for the NS operation performs better than a genie-aided, i.e. *better-than-optimal*, tractable detector for the S operation, then the same will hold for the corresponding optimal detectors.

We describe the approach for the FC case as the CC case follows in a similar fashion. For the FC-NS case the second equation of (7) is extended to $\mathbf{y}_t = \sqrt{\rho}\Phi_t\mathbf{h}s_t + \mathbf{z}_t$, for $t = 1, \dots, T$, where the m -th element of the diagonal matrix Φ_t is $e^{j\sum_{\tau=1}^t\phi_m[\tau]}$ and the increments $\phi_m[\tau]$ are i.i.d. distributed according to (3). In the following we provide a suboptimal causal SBS detector with decision feedback for the symbol s_t . At time t the information symbols up to $t - 1$ have been detected and this information is used by the detector as if the detected values were the true ones. Further, for simplicity the detector substitutes the random rotations of the phase noise rotations at time $1 \leq \tau \leq t - 1$ with their statistical mean $d_m[\tau] = \mathbb{E}\left[e^{j\sum_{\hat{\tau}=1}^{\tau}\phi_m[\hat{\tau}]}\right] = \alpha_{1,\tau}$, where we make the simplification that the phase noise statistics along the BS antennas are the same. Hence, the detector assumes that at some $\tau \in [1, t - 1]$ the received vector \mathbf{y}_τ is observed via $\mathbf{y}_\tau = \sqrt{\rho}\text{diag}\{d_1[\tau], \dots, d_M[\tau]\}\mathbf{h}\hat{s}_\tau + \mathbf{z}_\tau$. Finally, at time t the phase noise impairment is assumed to have the same statistics as the accumulated phase noise increments up to t , $\sum_{\tau=1}^t\phi_m[\tau]$. Hence the causal

decision feedback detector selects the $s_t \in \mathcal{S}$ that maximizes the likelihood $p(\mathbf{x}, \mathbf{y}_1, \dots, \mathbf{y}_t | \hat{s}_1, \dots, \hat{s}_{t-1}, s_t)$. The final result follows by similar steps as in Proposition 1, i.e.

$$\begin{aligned} \hat{s}_t = \arg \max_{s_t \in \mathcal{S}} & -M \ln(\hat{a}_t + \rho |s_t|^2) + \frac{\rho \left(\sum_{m=1}^M |\hat{v}_m[t]|^2 + |s_t|^2 \|\mathbf{y}_t\|^2 \right)}{\hat{a}_t + \rho |s_t|^2} \\ & + \sum_{m=1}^M \ln \left(\alpha_{0,t} I_0 \left(\frac{2\rho |\hat{\chi}_m[t]|}{\hat{a}_t + \rho |s_t|^2} \right) \right) \\ & + 2 \sum_{p=1}^{\infty} \alpha_{p,t} I_p \left(\frac{2\rho |\hat{\chi}_m[t]|}{\hat{a}_t + \rho |s_t|^2} \right) \cos(p(\arg(\hat{\chi}_m[t]))) \end{aligned} \quad (42)$$

where $\hat{a}_t \triangleq 1 + \rho + \rho \sum_{\tau=1}^{t-1} |d_m[\tau] \hat{s}_\tau|^2$, $\hat{v}_m[t] \triangleq x_m^* + \sum_{\tau=1}^{t-1} y_m^*[\tau] d_m[\tau] \hat{s}_\tau$, $\hat{\chi}_m[t] \triangleq \hat{v}_{m,t}^* y_m^*[t] s_t$. The sequence $\alpha_{0,t}, \alpha_{1,t}, \dots$ is the Fourier expansion of the pdf of $\sum_{\tau=1}^t \phi_m[\tau]$. As noted, this detector is suboptimal but implementable. Hence, the optimal causal detector for the FC-NS case must perform at least as good as this suboptimal detector.

In the following, we provide a genie-aided detector for the FC-S case. In this case, at time t the received vector is given by $\mathbf{y}_t = \sqrt{\rho} e^{j \sum_{\tau=1}^t \phi[\tau]} \mathbf{h} s_t + \mathbf{z}_t$. Assume that at time t , the receiver is aware of $\vartheta[t] \triangleq \sum_{\tau=1}^{t-1} \phi[\tau]$ and the true s_1, \dots, s_{t-1} . The causal ML detector selects the symbol $s_t \in \mathcal{S}$ that maximizes the likelihood $p(\mathbf{x}, \mathbf{y}_1, \dots, \mathbf{y}_t | s_1, \dots, s_t, \vartheta[t])$, i.e.

$$\begin{aligned} \hat{s}_t = \arg \max_{s_t \in \mathcal{S}} & -M \ln(a_t + \rho |s_t|^2) + \frac{\rho \left(\|\mathbf{v}_{t-1}\|^2 + |s_t|^2 \|\mathbf{y}_t\|^2 \right)}{a_t + \rho |s_t|^2} \\ & + \ln \left(\alpha_0 I_0 \left(\frac{2\rho |\chi_t|}{a_t + \rho |s_t|^2} \right) \right) \\ & + 2 \sum_{p=1}^{\infty} \alpha_p I_p \left(\frac{2\rho |\chi_t|}{a_t + \rho |s_t|^2} \right) \cos(p(\vartheta[t] - \arg(\chi_t))) \end{aligned} \quad (43)$$

where $a_t \triangleq 1 + \rho + \rho \sum_{\tau=1}^{t-1} |s_\tau|^2$, $\mathbf{v}_{t-1} \triangleq \mathbf{x} + \sum_{\tau=1}^{t-1} s_\tau^* e^{-j(\theta_\tau + \phi_\tau)} \mathbf{y}_\tau$ and $\chi_t \triangleq s_t^* \mathbf{v}_{t-1}^H \mathbf{y}_t$. This detector performs better than the actual causal ML detector for the synchronous operation, since it has the additional knowledge of the true prior symbols and the evolution of the phase noise process up to $t-1$. In Fig. 8 the SER of the two detectors in (43) and (42) is plotted for data interval length $T = 20$ and QPSK symbols. The phase noise increments

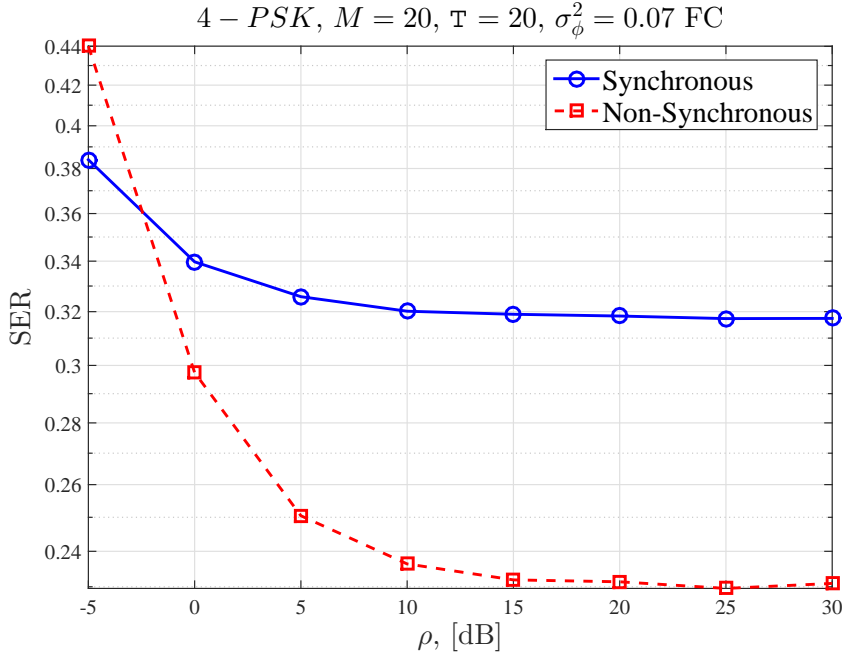


Figure 8: Comparison of the detectors (43) and (42) for $M = 20$ and $T = 20$.

$\phi_m[t]$ are assumed to be i.i.d. zero mean wrapped Gaussian with variance $\sigma_\phi^2 = 0.07$. This corresponds to the practical scenario of free-running oscillators. In this case, the pdf given by (3), where $\alpha_{m,p} = \left(\exp\left(-\frac{\sigma_\phi^2}{2}\right) \right)^{p^2}$ [28]. It is clear that the suboptimal FC-NS detector performs better than the “better-than-optimal” FC-S detector, which establishes the claim that the results of the previous sections are valid in more complex setups.

6 Conclusions

The problem of ML detection in a training-assisted single-user SIMO channel with phase noise impairments and M BS antennas was studied. A simple transmission protocol was considered, where one time slot is used for pilot and the second time slot is used for data transmission. The study included the case of constant and fading channel assumptions. For both assumptions on the channel gains two operations were investigated, i.e., the

synchronous and non-synchronous operations. Closed-form expressions of the optimal detectors were given for a general parameterization of the phase noise increments. SER floors were observed for all cases under study. For the synchronous operation the SER floors were independent of M for both the constant and fading channel case. In contrast the SER floor in the non-synchronous operation could be made arbitrarily small by increasing M .

The effects discovered in the paper are fundamental because they assume optimal detection in the Bayesian sense, as the detector was found by analytical marginalization of all nuisance parameters, i.e., the phase noise in the CC case and the channel fading and the phase noise in the FC case. That is, no filtering algorithm, alternative or additional, can improve the SER performance of any of the operations, S and NS. The observed SER floors and the distinction between S and NS operation are, thus, fundamental and not an artifact of suboptimal receiver processing. Further, we have shown that our results remain valid if the data interval is extended to multiple channel uses.

Appendices

7 Two-Slot Proofs

7.1 Proof of Proposition 1 for the CC-NS Case

Define the vectors $\boldsymbol{\theta} \triangleq [\theta_1, \dots, \theta_M]^T$ and $\boldsymbol{\phi} \triangleq [\phi_1, \dots, \phi_M]^T$. The likelihood function is given by

$$\begin{aligned}
 p(\mathbf{x}, \mathbf{y}|s) &= \iint p(\mathbf{x}, \mathbf{y}|s, \boldsymbol{\theta}, \boldsymbol{\phi})p(\boldsymbol{\theta}, \boldsymbol{\phi}|s)d\boldsymbol{\theta}d\boldsymbol{\phi} \\
 &\stackrel{(a)}{=} \iint p(\mathbf{x}|\boldsymbol{\theta})p(\mathbf{y}|s, \boldsymbol{\theta}, \boldsymbol{\phi})p(\boldsymbol{\theta})p(\boldsymbol{\phi})d\boldsymbol{\theta}d\boldsymbol{\phi} \\
 &\stackrel{(b)}{=} \prod_{m=1}^M \int_{-\pi}^{\pi} p(x_m|\theta_m)p(\theta_m) \underbrace{\int_{-\pi}^{\pi} p(y_m|s, \theta_m, \phi_m)p(\phi_m)d\phi_m}_{\triangleq p(y_m|\theta_m, s)} d\theta_m \\
 &= \prod_{n=1}^M p(x_n, y_n|s), \tag{44}
 \end{aligned}$$

where (a) follows from the fact that conditioned on $\boldsymbol{\theta}$, $\boldsymbol{\phi}$ and s , the vectors \mathbf{x} and \mathbf{y} are independent and that $\boldsymbol{\theta}$, $\boldsymbol{\phi}$ are independent of each other and s . The equality in (b) is a consequence of the independence of the components in $\boldsymbol{\theta}$, $\boldsymbol{\phi}$, \mathbf{w} and \mathbf{z} . The channel probability law $p(y_m|s, \theta_m, \phi_m)$ can be

expressed as

$$\begin{aligned}
 p(y_m|s, \theta_m, \phi_m) &= \frac{1}{\pi} e^{-|y_m - \sqrt{\rho} g_m e^{j(\theta_m + \phi_m)} s|^2} \\
 &= \frac{1}{\pi} e^{-|y_m|^2 - \rho g_m^2 |s|^2 + 2\sqrt{\rho} g_m |s^* y_m| \cos(\phi_m + \theta_m - \arg(s^* y_m))} \\
 &= \frac{e^{-|y_m|^2 - \rho g_m^2 |s|^2}}{\pi} \left(I_0(2\sqrt{\rho} g_m |s^* y_m|) \right. \\
 &\quad \left. + 2 \sum_{l=1}^{\infty} I_l(2\sqrt{\rho} g_m |s^* y_m|) \cos(l(\phi_m + \theta_m - \arg(s^* y_m))) \right), \tag{45}
 \end{aligned}$$

where the last step follows from the Jacobi-Anger formula [39, Section 4.4, p. 100]

$$e^{\alpha \cos \beta} = I_0(\alpha) + 2 \sum_{l=1}^{\infty} I_l(\alpha) \cos(l\beta). \tag{46}$$

Then, the conditional $p(y_m|\theta_m, s)$ is

$$\begin{aligned}
 p(y_m|\theta_m, s) &= \int_{-\pi}^{\pi} p(y_m|s, \theta_m, \phi_m) p(\phi_m) d\phi_m \\
 &= \frac{e^{-|y_m|^2 - \rho g_m^2 |s|^2}}{\pi} \left(\alpha_{m,0} I_0(2\sqrt{\rho} g_m |s^* y_m|) \right. \\
 &\quad \left. + 2 \sum_{l=1}^{\infty} \alpha_{m,l} I_l(2\sqrt{\rho} g_m |s^* y_m|) \cos(l(\theta_m - \arg(s^* y_m))) \right).
 \end{aligned}$$

By manipulating $p(x_m|\theta_m)$ in the same way as in (45) and by the orthogonality of cos and sin we obtain

$$\begin{aligned}
 p(x_m, y_m|s) &= \int_{-\pi}^{\pi} p(x_m|\theta_m) p(\theta_m) p(y_m|\theta_m, s) d\theta_m \\
 &= \frac{e^{-|x_m|^2 - |y_m|^2 - \rho g_m^2 (1+|s|^2)}}{\pi^2} \left(\beta_{m,0} + 2 \sum_{l=1}^{\infty} \beta_{m,l} \cos(l\zeta_m) \right). \tag{47}
 \end{aligned}$$

The result in (9), (10), (11) and (12) follows by substituting (47) in (44).

7.2 Proof of Proposition 1 for the FC-NS Case

The likelihood function in this case is given by

$$\begin{aligned}
p(\mathbf{x}, \mathbf{y}|s) &= \int p(\mathbf{x}, \mathbf{y}|s, \phi, \mathbf{h})p(\mathbf{h}, \phi|s)d\mathbf{h}d\phi \\
&\stackrel{(a)}{=} \int p(\mathbf{x}|\mathbf{h})p(\mathbf{y}|s, \phi, \mathbf{h})p(\mathbf{h})p(\phi)d\mathbf{h}d\phi \\
&= \int p(\phi) \int p(\mathbf{h})p(\mathbf{y}|s, \phi, \mathbf{h})p(\mathbf{x}|\mathbf{h})d\mathbf{h}d\phi \\
&\stackrel{(b)}{=} \prod_{m=1}^M \int_{-\pi}^{\pi} p(\phi_m) \underbrace{\int_{\mathbb{C}} p(h_m)p(y_m|s, \phi_m, h_m)p(x_m|h_m)dh_m}_{\triangleq p(x_m, y_m|s, \phi_m)} d\phi_m, \quad (48)
\end{aligned}$$

where (a) follows from the fact that conditioned on s , \mathbf{h} and ϕ , the received vectors \mathbf{x} and \mathbf{y} are independent and from the assumption that s , \mathbf{h} and ϕ are mutually independent. Further, the factorization in (b) follows from the fact that the components of \mathbf{h} and ϕ are independent. It holds that

$$p(x_m|h_m) = \frac{e^{-|x_m - \sqrt{\rho}h_m|^2}}{\pi} = \frac{e^{-|x_m|^2 - \rho|h_m|^2}}{\pi} \exp(\Re\{2\sqrt{\rho}x_m^*h_m\}), \quad (49)$$

$$\begin{aligned}
p(y_m|\phi_m, h_m, s) &= \frac{e^{-|y_m - \sqrt{\rho}e^{j\phi_m}h_ms|^2}}{\pi} \\
&= \frac{e^{-|y_m|^2 - \rho|s|^2|h_m|^2}}{\pi} \exp\left(\Re\left\{2\sqrt{\rho}y_m^*h_ms e^{j\phi_m}\right\}\right). \quad (50)
\end{aligned}$$

From (49), (50) and $p(h_m) = \frac{1}{\pi}e^{-|h_m|^2}$ the conditional pdf $p(x_m, y_m|s, \phi_m)$ is written as

$$\begin{aligned}
p(x_m, y_m|s, \phi_m) &= \int_{\mathbb{C}} p(h_m)p(y_m|s, \phi_m, h_m)p(x_m|h_m)dh_m = \frac{e^{-|x_m|^2 - |y_m|^2}}{\pi^3} \\
&\times \int_{\mathbb{C}} \exp\left(-(\rho + \rho|s|^2 + 1) \left[|h_m|^2 - 2\Re\left\{\frac{\sqrt{\rho}(x_m + s^*y_me^{-j\phi_m})^*h_m}{\rho + \rho|s|^2 + 1}\right\}\right]\right) dh_m
\end{aligned}$$

For notational convenience we define $c_m \triangleq \frac{\sqrt{\rho}(x_m + s^* y_m e^{-j\phi_m})}{\rho + \rho|s|^2 + 1}$ and $p(x_m, y_m | s, \phi_m)$ is expressed as

$$\begin{aligned} p(x_m, y_m | s, \phi_m) &= \frac{e^{-|x_m|^2 - |y_m|^2}}{\pi^3} \int_{\mathbb{C}} e^{-(\rho + \rho|s|^2 + 1)[|h_m|^2 - 2\Re\{c_m^* h_m\} + |c_m|^2 - |c_m|^2]} dh_m \\ &\stackrel{(a)}{=} \frac{e^{-|x_m|^2 - |y_m|^2 + (\rho + \rho|s|^2 + 1)|c_m|^2}}{\pi^3} \int_{\mathbb{C}} e^{-(\rho + \rho|s|^2 + 1)|h_m - c_m|^2} dh_m \\ &\stackrel{(b)}{=} \frac{e^{-|x_m|^2 - |y_m|^2 + (\rho + \rho|s|^2 + 1)|c_m|^2}}{\pi^2 (\rho + \rho|s|^2 + 1)}, \end{aligned}$$

where in (a) we complete the square and (b) is the integral of a complex Gaussian function over \mathbb{C} . By using these expressions we can compute

$$\begin{aligned} p(x_m, y_m | s) &= \int_{-\pi}^{\pi} p(\phi_m) p(x_m, y_m | s, \phi_m) d\phi_m \\ &= \frac{1}{\pi^2 (\rho + \rho|s|^2 + 1)} e^{-\frac{1 + \rho|s|^2}{\rho + \rho|s|^2 + 1}|x_m|^2 - \frac{1 + \rho}{\rho + \rho|s|^2 + 1}|y_m|^2} \\ &\quad \times \int_{-\pi}^{\pi} p(\phi_m) e^{\frac{2\rho|s^* x_m^* y_m|}{\rho + \rho|s|^2 + 1} \cos(\arg(x_m^* y_m) - \phi_m - \arg(s))} d\phi_m. \end{aligned} \quad (51)$$

For convenience of the notation, we define

$$\begin{aligned} A_{m,1} &\triangleq \frac{e^{-\frac{1 + \rho|s|^2}{\rho + \rho|s|^2 + 1}|x_m|^2 - \frac{1 + \rho}{\rho + \rho|s|^2 + 1}|y_m|^2}}{\pi^2 (\rho + \rho|s|^2 + 1)}, \quad A_{m,2} \triangleq \frac{2\rho|s^* x_m^* y_m|}{\rho + \rho|s|^2 + 1}, \\ \text{and } A_{m,3} &\triangleq \arg(s) - \arg(x_m^* y_m). \end{aligned} \quad (52)$$

By substituting (3) and (52) into (51)

$$\begin{aligned} p(x_m, y_m | s) &= \frac{A_{m,1}}{2\pi} \int_{-\pi}^{\pi} \left(\alpha_{m,0} + 2 \sum_{l=1}^{\infty} \alpha_{m,l} \cos(l\phi_m) \right) e^{A_{m,2} \cos(\phi_m + A_{m,3})} d\phi_m \\ &= \frac{A_{m,1} \alpha_{m,0}}{2\pi} \int_{-\pi}^{\pi} e^{A_{m,2} \cos(\phi_m + A_{m,3})} d\phi_m \\ &\quad + \frac{A_{m,1}}{2\pi} \sum_{l=1}^{\infty} \int_{-\pi}^{\pi} 2\alpha_{m,l} \cos(l\phi_m) e^{A_{m,2} \cos(\phi_m + A_{m,3})} d\phi_m = A_{m,1} \alpha_{m,0} I_0(A_{m,2}) \\ &\quad + \frac{A_{m,1}}{\pi} \sum_{l=1}^{\infty} \alpha_{m,l} \int_{-\pi}^{\pi} \cos(l\phi_m) \left(I_0(A_{m,2}) + 2 \sum_{k=1}^{\infty} I_k(A_{m,2}) \cos(k(\phi_m + A_{m,3})) \right) d\phi_m \\ &= A_{m,1} \left(\alpha_{m,0} I_0(A_{m,2}) + 2 \sum_{l=1}^{\infty} \alpha_{m,l} I_l(A_{m,2}) \cos(lA_{m,3}) \right). \end{aligned} \quad (53)$$

Since $p(x_m, y_m|s)$ is a density function and hence it is absolutely integrable, the interchange of summation and integration is possible by the Fubini-Tonelli theorem [37]. The result in (9), (13), (14) and (12) follows by substituting (52) and (53) into (48).

7.3 Proof of Proposition 2 for the CC-S Case

The likelihood function in this case is factorized similarly to (44) as

$$p(\mathbf{x}, \mathbf{y}|s) = \int p(\mathbf{x}|\theta)p(\theta) \int p(\mathbf{y}|s, \theta, \phi)p(\phi)d\phi d\theta. \quad (54)$$

The integrals in (54) can be calculated using a similar approach as in Appendix 7.1 yielding (20), (10), (21) and (22).

7.4 Proof of Proposition 2 for the FC-S Case

The likelihood function in this case is factorized similarly to (48) as

$$p(\mathbf{x}, \mathbf{y}|s) = \int_{-\pi}^{\pi} p(\phi) \int_{\mathbb{C}^M} p(\mathbf{y}|s, \phi, \mathbf{h})p(\mathbf{h})p(\mathbf{x}|\mathbf{h})d\mathbf{h}d\phi. \quad (55)$$

The integrals in (55) can be calculated using a similar approach as in Appendix 7.2 yielding (20), (13), (23) and (24).

7.5 Proof of Proposition 3

The proof is similar to the proof in Appendix 7.2. We substitute the von Mises pdf $p(\phi_m) = \frac{e^{\kappa \cos \phi_m}}{2\pi I_0(\kappa)}$ into (51) and evaluate the resulting integral. This yields (29). The proof for (30) is similar to the proof for (29).

8 T-slot Detectors

In this appendix detectors are presented for the T-data-slot extended transmission protocol. Due to the fact that the derivation of the optimal detectors in the maximum-likelihood (ML) sense in closed form appears to be intractable, suboptimal detectors for the non-synchronous (NS) operation and genie-aided, i.e. *better-than-optimal*, detectors for the synchronous (S) operation are derived. By simulations it is shown that the performance of the suboptimal NS detectors is superior to the genie-aided S detectors. This conclusively establishes the claim that the results of the two-slot model are valid for the extended models as well.

8.1 Decision Feedback Detector: CC-NS

We consider a transmission interval of $T + 1$ channel uses. Training is done during the first channel use and uncoded information symbols are transmitted during the subsequent T channel uses. The received vector during training, \mathbf{x} , and the received vectors during data transmission, \mathbf{y}_t , at time t are given by

$$\mathbf{x} = \sqrt{\rho}\mathbf{\Theta}\mathbf{g} + \mathbf{w}, \quad (56)$$

$$\mathbf{y}_t = \sqrt{\rho}\mathbf{\Theta}\mathbf{\Phi}_t\mathbf{g}s_t + \mathbf{z}_t, \quad t = 1, \dots, T, \quad (57)$$

where $\mathbf{\Theta} \triangleq \text{diag}\{e^{j\theta_1}, \dots, e^{j\theta_M}\}$ and $\mathbf{\Phi}_t \triangleq \text{diag}\{e^{j\sum_{\tau=1}^t\phi_1[\tau]}, \dots, e^{j\sum_{\tau=1}^t\phi_M[\tau]}\}$. The phases θ_m , $m = 1, \dots, M$ are random i.i.d. initial phases uniformly distributed in $[-\pi, \pi)$. The phase noise increments $\phi_m[t]$ are i.i.d random variables with pdf parameterized by

$$p_{\Phi_m[\tau]}(\phi_m[\tau]) = \frac{1}{2\pi} \left(\alpha_{m,0} + 2 \sum_{p=1}^{\infty} \alpha_{m,p} \cos(p\phi_m[\tau]) \right). \quad (58)$$

The vector $\mathbf{g} \in \mathbb{R}^M$ is a constant and known vector of amplitudes and s_t are data symbols selected from a fixed constellation \mathcal{S} . In order to derive a suboptimal detector for the symbol s_t we consider that the receiver uses the detected symbols $\hat{s}_1, \dots, \hat{s}_{t-1}$ as if they were true. Further, the exact knowledge of the phase noise increments is not available and can be only

estimated. For simplicity of the detector we assume that distortion due to phase noise is equal to the expected value of it conditioned on the matrix of the initial phases Θ . Hence, the received vectors \mathbf{y}_τ , $1 \leq \tau \leq t-1$ are assumed to be observed by

$$\mathbf{y}_\tau = \sqrt{\rho} \Theta \mathbf{D}_\tau \mathbf{g} \hat{s}_\tau + \mathbf{z}_\tau, \quad 1 \leq \tau \leq t-1, \quad (59)$$

where $\mathbf{D}_\tau \triangleq \mathbb{E}[\Phi_\tau] \triangleq \text{diag}\{\alpha_{1,1,\tau}, \dots, \alpha_{M,1,\tau}\}$, where $\alpha_{m,1,\tau}$ are the coefficients of the Fourier expansion of the pdf of $\sum_{\hat{\tau}=1}^{\tau} \phi_m[\hat{\tau}]$. Finally, at time t the received vector is given by (57), where the phase noise perturbation equals to the total accumulated phase noise up to time t , $\varphi_m[t] \triangleq \sum_{\tau=1}^t \phi_m[\tau]$. The suboptimal causal decision feedback receiver is given by

$$\hat{s}_t = \arg \max_{s_t \in \mathcal{S}} p(\mathbf{x}, \mathbf{y}_1^t | \hat{s}_1^{t-1}, s_t), \quad (60)$$

where we denote the sequence $v_a^b \triangleq \{v_a, \dots, v_b\}$. The detailed steps of the derivation follow. The likelihood $p(\mathbf{x}, \mathbf{y}_1^t | \hat{s}_1^{t-1}, s_t)$ is written as

$$\begin{aligned} p(\mathbf{x}, \mathbf{y}_1^t | \hat{s}_1^{t-1}, s_t) &= \int \cdots \int p(\mathbf{x}, \mathbf{y}_1^t | \hat{s}_1^{t-1}, s_t, \Theta, \Phi_t) p(\Theta) p(\Phi_t) d\Theta d\Phi_t \\ &= \prod_{m=1}^M \iint p(x_m, \{y_m[\tau]\}_1^t | \hat{s}_1^{t-1}, s_t, \theta_m, \varphi_m[t]) p(\theta_m) p(\varphi_m[t]) d\theta_m d\varphi_m[t]. \end{aligned}$$

The derivation proceeds with marginalizing over the accumulated phase noise $\varphi_m[t]$.

$$p(x_m, \{y_m[\tau]\}_1^t | \hat{s}_1^{t-1}, s_t, \theta_m) = \int_{-\pi}^{\pi} p(x_m, \{y_m[\tau]\}_1^t | \hat{s}_1^{t-1}, s_t, \theta_m, \varphi_m[t]) p(\varphi_m[t]) d\varphi_m[t],$$

where

$$p(\varphi_m[t]) = \frac{1}{2\pi} \left(\alpha_{0,m,t} + 2 \sum_{p=1}^{\infty} \alpha_{p,m,t} \cos(p\varphi_m[t]) \right).$$

Note that here the coefficients of the Fourier expansion of $p(\varphi_m[t])$ depend on t , since the larger the t the larger the variance of $\varphi_m[t]$. The density $p(x_m, \{y_m[\tau]\}_1^t | \hat{s}_1^{t-1}, s_t, \theta_m, \varphi_m[t])$ is given by

$$\begin{aligned} p(x_m, \{y_m[\tau]\}_1^t | \hat{s}_1^{t-1}, s_t, \theta_m, \varphi_m[t]) &= \frac{1}{\pi^{t+1}} e^{-|x_m - \sqrt{\rho} e^{j\theta_m} g_m|^2 - |y_m[t] - \sqrt{\rho} e^{j\theta_m} e^{j\varphi_m[t]} g_m s_t|^2} \\ &\quad \times e^{-\sum_{\tau=1}^{t-1} |y_m[\tau] - \sqrt{\rho} e^{j\theta_m} d_m[\tau] g_m \hat{s}_\tau|^2}. \end{aligned}$$

With the use of the Jacobi-Anger formula, the phase noise term $\varphi_m[t]$ can be marginalized out and the density $p(x_m, \{y_m[\tau]\}_1^t | \hat{s}_1^{t-1}, s_t, \theta_m)$ is given by

$$\begin{aligned} p(x_m, \{y_m\}_1^t | \hat{s}_1^{t-1}, s_t, \theta_m) &= \frac{1}{\pi^{t+1}} e^{-|x_m|^2 - \sum_{\tau=1}^t |y_m[\tau]|^2 - \rho g_m^2 (b_m[t] + |s_t|^2)} \\ &\times e^{2\sqrt{\rho} |c_m[t] g_m| \cos(\theta_m + \arg(c_m[t] g_m))} \left(\alpha_{0,m,t} I_0(2\sqrt{\rho} |B_m[t]|) \right. \\ &\left. + 2 \sum_{q=1}^{\infty} \alpha_{q,m,t} I_q(2\sqrt{\rho} |B_m[t]|) \cos(q(\theta_m + \arg(B_m[t]))) \right), \end{aligned}$$

where $B_m[t] \triangleq y_m^*[t] g_m s_t$, $b_m[t] \triangleq 1 + \sum_{\tau=1}^{t-1} |d_m[\tau] \hat{s}_\tau|^2$, $c_m[t] \triangleq x_m^* + \sum_{\tau=1}^{t-1} y_m^*[\tau] d_m[\tau] \hat{s}_\tau$. Further marginalization with respect to θ_m can be carried out which yields the density

$$\begin{aligned} p(x_m, \{y_m[\tau]\}_1^t | \hat{s}_1^{t-1}, s_t) &= \int_{-\pi}^{\pi} p(x_m, \{y_m[\tau]\}_0^t | \hat{s}_1^{t-1}, s_t, \theta_m) p(\theta_m) d\theta_m \\ &= \frac{1}{\pi^{t+1}} e^{-|x_m|^2 - \sum_{\tau=1}^t |y_m[\tau]|^2 - \rho g_m^2 (b_m[t] + |s_t|^2)} \left(\alpha_{0,m,t} I_0(2\sqrt{\rho} |B_m[t]|) I_0(2\sqrt{\rho} |c_m[t] g_m|) \right. \\ &\left. + 2 \sum_{p=1}^{\infty} \alpha_{p,m,t} I_p(2\sqrt{\rho} |B_m[t]|) I_p(2\sqrt{\rho} |c_m[t] g_m|) \cos(p(\arg(B_m[t]) - \arg(c_m[t] g_m))) \right). \end{aligned}$$

The detector is finally given by

$$\begin{aligned} \hat{s}_t &= \arg \max_{s_t \in \mathcal{S}} -\rho |s_t|^2 \|g\|^2 + \sum_{m=1}^M \ln \left(\alpha_{0,m,t} I_0(2\sqrt{\rho} |B_m[t]|) I_0(2\sqrt{\rho} |c_m[t] g_m|) \right) \quad (61) \\ &+ 2 \sum_{p=1}^{\infty} \alpha_{p,m,t} I_p(2\sqrt{\rho} |B_m[t]|) I_p(2\sqrt{\rho} |c_m[t] g_m|) \cos(p(\arg(B_m[t]) - \arg(c_m[t] g_m))) \Big). \end{aligned}$$

8.2 “Better-than-optimal” Detector: CC-S

The system model in this case follows trivially from (56) and (57) by setting $\theta_m \equiv \theta$ and $\phi_m[t] \equiv \phi_t, \forall m \in \{1, \dots, M\}$. Hence, the system model at time t is given by

$$\mathbf{x} = \sqrt{\rho} e^{j\theta} \mathbf{g} + \mathbf{w}, \quad (62)$$

$$\mathbf{y}_t = \sqrt{\rho} e^{j(\theta_t + \phi_t)} \mathbf{g} s_t + \mathbf{z}_t, \quad t = 1, \dots, T, \quad (63)$$

where $\theta_t \triangleq \theta + \sum_{\tau=1}^{t-1} \phi_\tau$. The likelihood for the causal ML symbol-by-symbol (SBS) detector for the symbol s_t is given by

$$p(\mathbf{x}, \mathbf{y}_1^t | s_t) = \sum_{\substack{s_\tau \in \mathcal{S}, \\ \tau \in 1, \dots, t-1}} \int \cdots \int_{-\pi}^{\pi} p(\mathbf{x}, \mathbf{y}_1^t | \theta_0, \phi_1^t, s_1^t) p(\theta_0) \prod_{\tau=1}^t p(\phi_\tau) d\theta_0 d\phi_1^t. \quad (64)$$

The derivation of a closed-form expression of the above decision rule appears to be mathematically intractable. Hence, we are interested in deriving a genie-aided, "better-than-optimal" detector, that will yield an optimistic performance bound on the detector in (64). Assume that before the detection of the symbol s_t , a genie provides the receiver with the exact knowledge of θ_t . Then, the likelihood for ML causal SBS detector for s_t is

$$p(\mathbf{y}_t | \theta_t, s_t) = \int_{-\pi}^{\pi} p(\mathbf{y}_t | \theta_t, \phi_t, s_t) p(\phi_t) d\phi_t, \quad (65)$$

where

$$\begin{aligned} p(\mathbf{y}_t | \theta_t, \phi_t, s_t) &= \frac{1}{\pi^M} e^{-\|\mathbf{y}_t - \sqrt{\rho} e^{j(\theta_t + \phi_t)} \mathbf{g} s_t\|^2} \\ &= \frac{e^{-\|\mathbf{y}_t\|^2 - \rho |s_t|^2 \|\mathbf{g}\|^2}}{\pi^M} e^{2\sqrt{\rho} |\mathbf{y}_t^H \mathbf{g} s_t| \cos(\phi_t + \theta_t + \arg(s_t \mathbf{y}_t^H \mathbf{g}))}. \end{aligned} \quad (66)$$

From (58) and (66) the likelihood in (65) is given by

$$p(\mathbf{y}_t | \theta_t, s_t) = \frac{e^{-\|\mathbf{y}_t\|^2 - \rho |s_t|^2 \|\mathbf{g}\|^2}}{\pi^M} \left(\beta_0 + 2 \sum_{p=1}^{\infty} \beta_p \cos(p(\theta_t + \arg(s_t \mathbf{y}_t^H \mathbf{g}))) \right), \quad (67)$$

where $\beta_p \triangleq \alpha_p I_p(2\sqrt{\rho} |\mathbf{y}_t^H \mathbf{g} s_t|)$. Finally the detector is given by

$$\hat{s}_t = \arg \max_{s_\tau \in \mathcal{S}} -\rho |s_t|^2 \|\mathbf{g}\|^2 + \ln \left(\beta_0 + 2 \sum_{p=1}^{\infty} \beta_p \cos(p(\theta_t + \arg(s_t \mathbf{y}_t^H \mathbf{g}))) \right). \quad (68)$$

8.3 Decision Feedback Detector: FC-NS

The system model for the fading channel case is given by

$$\mathbf{x} = \sqrt{\rho} \mathbf{h} + \mathbf{w}, \quad (69)$$

$$\mathbf{y}_t = \sqrt{\rho} \Phi_t \mathbf{h} s_t + \mathbf{z}_t, \quad t = 1, \dots, T, \quad (70)$$

where the difference with (56) and (57) is the fading channel $\mathbf{h} \sim \mathcal{N}_C(\mathbf{0}, \mathbf{I}_M)$. The initial phase reference Θ has been absorbed into \mathbf{h} without modifying its statistics. The causal SBS decision feedback detector is derived also similarly to the constant channel case by assuming the following auxiliary model

$$\begin{aligned} \mathbf{x} &= \sqrt{\rho}\mathbf{h} + \mathbf{w}, \\ \mathbf{y}_\tau &= \sqrt{\rho}\mathbf{D}_\tau\mathbf{h}\hat{s}_\tau + \mathbf{z}_\tau, \quad \tau = 1, \dots, t-1 \\ \mathbf{y}_t &= \sqrt{\rho}\Phi_t\mathbf{h}s_t + \mathbf{z}_t. \end{aligned}$$

The decision rule for this detector is given by (60). The density $p(\mathbf{x}, \mathbf{y}_1^t | \hat{s}_1^{t-1}, s_t)$ in this case is given by

$$\begin{aligned} p(\mathbf{x}, \mathbf{y}_1^t | \hat{s}_1^{t-1}, s_t) &= \int \cdots \int p(\mathbf{x}, \mathbf{y}_1^t | \hat{s}_1^{t-1}, s_t, \Phi_t, \mathbf{h}) p(\mathbf{h}) p(\Phi_t) d\mathbf{h} d\Phi_t \\ &= \prod_{m=1}^M \iint p(x_m, \{y_m[\tau]\}_1^t | \hat{s}_1^{t-1}, s_t, \varphi_m[t], h_m) p(h_m) p(\varphi_m[t]) d\varphi_m[t] dh_m. \end{aligned}$$

At the first step of the derivation the channel h_m is marginalized out which yields the result for the density $p(x_m, \{y_m[\tau]\}_1^t | \hat{s}_1^{t-1}, s_t, \varphi_m[t])$

$$\begin{aligned} p(x_m, \{y_m[\tau]\}_1^t | \hat{s}_1^{t-1}, s_t, \varphi_m[t]) &= \frac{e^{-|x_m|^2 - \sum_{\tau=1}^t |y_m[\tau]|^2 + \frac{\rho(|\hat{v}_m[t]|^2 + |y_m[t]|^2 |s_t|^2)}{\hat{a}_m[t] + \rho|s_t|^2}}}{\pi^{t+1} (\hat{a}_m[t] + \rho|s_t|^2)} \\ &\cdot \left(I_0 \left(\frac{2\rho|\hat{\chi}_m[t]|}{\hat{a}_m[t] + \rho|s_t|^2} \right) + 2 \sum_{p=1}^{\infty} I_p \left(\frac{2\rho|\hat{\chi}_m[t]|}{\hat{a}_m[t] + \rho|s_t|^2} \right) \cos(p(\varphi_m[t] + \arg(\hat{\chi}_m[t]))) \right), \end{aligned}$$

where $\hat{a}_m[t] = 1 + \rho + \rho \sum_{\tau=1}^{t-1} |d_m[\tau]\hat{s}_\tau|^2$, $\hat{v}_m[t] \triangleq x_m^* + \sum_{\tau=1}^{t-1} y_m^*[\tau]d_m[\tau]\hat{s}_\tau$ and $\hat{\chi}_m[t] \triangleq \hat{v}_m^*[t]y_m^*[t]s_t$. In the following, the variable $\varphi_m[t]$ is marginalized out.

$$\begin{aligned} p(x_m, \{y_m[\tau]\}_1^t | \hat{s}_1^{t-1}, s_t) &= \int p(x_m, \{y_m[\tau]\}_1^t | \hat{s}_1^{t-1}, s_t, \varphi_m[t]) p(\varphi_m[t]) d\varphi_m[t] \\ &= \frac{1}{\pi^{t+1} (\hat{a}_m[t] + \rho|s_t|^2)} e^{-|x_m|^2 - \sum_{\tau=1}^t |y_m[\tau]|^2 + \frac{\rho(|\hat{v}_m[t]|^2 + |y_m[t]|^2 |s_t|^2)}{\hat{a}_m[t] + \rho|s_t|^2}} \\ &\times \left(\alpha_{0,m,t} I_0 \left(\frac{2\rho|\hat{\chi}_m[t]|}{\hat{a}_m[t] + \rho|s_t|^2} \right) + 2 \sum_{p=1}^{\infty} \alpha_{p,m,t} I_p \left(\frac{2\rho|\hat{\chi}_m[t]|}{\hat{a}_m[t] + \rho|s_t|^2} \right) \cos(p(\arg(\hat{\chi}_m[t]))) \right). \end{aligned}$$

Finally, detector is given by

$$\hat{s}_t = \arg \max_{s \in \mathcal{S}} - \sum_{m=1}^M \ln(\hat{a}_m[t] + \rho|s_t|^2) + \sum_{m=1}^M \frac{\rho(|\hat{v}_m[t]|^2 + |\mathbf{y}_m[t]|^2|s_t|^2)}{\hat{a}_m[t] + \rho|s_t|^2} \quad (71)$$

$$+ \sum_{m=1}^M \ln\left(\alpha_{0,m,t} I_0\left(\frac{2\rho|\hat{\chi}_m[t]|}{\hat{a}_t + \rho|s_t|^2}\right) + 2 \sum_{p=1}^{\infty} \alpha_{p,m,t} I_p\left(\frac{2\rho|\hat{\chi}_m[t]|}{\hat{a}_t + \rho|s_t|^2}\right) \cos(p(\arg(\hat{\chi}_m[t])))\right).$$

8.4 “Better-than-optimal” Detector: FC-S

The system model is given by

$$\mathbf{x} = \sqrt{\rho}\mathbf{h} + \mathbf{w}, \quad (72)$$

$$\mathbf{y}_t = \sqrt{\rho}e^{j(\sum_{\tau=1}^{t-1} \phi_{\tau} + \phi_t)} \mathbf{h} s_t + \mathbf{z}_t, \quad t = 1, \dots, T. \quad (73)$$

The likelihood for the causal ML symbol-by-symbol (SBS) detector for the symbol s_t is given by

$$p(\mathbf{x}, \mathbf{y}_1^t | s_t) = \sum_{\substack{s_{\tau} \in \mathcal{S}, \\ \tau=1, \dots, t-1}} \int \cdots \int_{-\pi}^{\pi} p(\mathbf{x}, \mathbf{y}_1^t | \mathbf{h}, \phi_1^t, s_1^t) p(\mathbf{h}) \prod_{\tau=1}^t p(\phi_{\tau}) d\mathbf{h} d\phi_{\tau}. \quad (74)$$

We derive a genie-aided, “better-than-optimal” detector, that will yield an optimistic performance bound on the detector in (74). Assume that before the detection of the symbol s_t , a genie provides the receiver with the exact knowledge of $\sum_{\tau=1}^{t-1} \phi_{\tau}$ and the previously transmitted symbols, s_1^{t-1} . Then, the likelihood for ML SBS detector for s_t is

$$p(\mathbf{x}, \mathbf{y}_1^t | \phi_1^{t-1}, s_1^{t-1}, s_t) = \int_{-\pi}^{\pi} \int_{\mathbb{C}^M} p(\mathbf{x}, \mathbf{y}_1^t | \mathbf{h}, \phi_1^{t-1}, \phi_t, s_1^{t-1}, s_t) p(\mathbf{h}) p(\phi_t) d\mathbf{h} d\phi_t \quad (75)$$

The conditional density $p(\mathbf{x}, \mathbf{y}_1^t | \mathbf{h}, \phi_1^{t-1}, \phi_t, s_1^{t-1}, s_t)$ is given by

$$p(\mathbf{x}, \mathbf{y}_1^t | \mathbf{h}, \phi_1^{t-1}, \phi_t, s_1^{t-1}, s_t) = \frac{1}{\pi^{M(t+1)}} e^{-\|\mathbf{x} - \sqrt{\rho}\mathbf{h}\|^2 - \sum_{\tau=1}^t \|\mathbf{y}_{\tau} - \sqrt{\rho}\mathbf{h}e^{j(\sum_{\hat{\tau}=1}^{\tau} \phi_{\hat{\tau}} + \phi_{\tau})} s_{\tau}\|^2}.$$

Define $a_t \triangleq 1 + \rho + \rho \sum_{\tau=1}^{t-1} |s_{\tau}|^2$ and $\mathbf{v}_t \triangleq \mathbf{x} + \sum_{\tau=1}^t s_{\tau}^* e^{-j(\sum_{\hat{\tau}=1}^{\tau} \phi_{\hat{\tau}})} \mathbf{y}_{\tau}$. The density $p(\mathbf{x}, \mathbf{y}_1^t | \phi_1^{t-1}, \phi_t, s_1^{t-1}, s_t)$ follows directly

$$p(\mathbf{x}, \mathbf{y}_1^t | \phi_1^{t-1}, \phi_t, s_1^{t-1}, s_t) = \int_{\mathbb{C}^M} p(\mathbf{x}, \mathbf{y}_1^t | \mathbf{h}, \phi_1^{t-1}, \phi_t, s_1^{t-1}, s_t) p(\mathbf{h}) d\mathbf{h}$$

$$= \frac{e^{-\|\mathbf{x}\|^2 - \sum_{\tau=1}^t \|\mathbf{y}_{\tau}\|^2 + \frac{\rho(\|\mathbf{v}_{t-1}\|^2 + |s_t|^2 \|\mathbf{y}_t\|^2)}{a_t + \rho|s_t|^2} + \frac{2\rho|s_t^* \mathbf{v}_{t-1}^H \mathbf{y}_t| \cos(\phi_t + \sum_{\tau=1}^{t-1} \phi_{\tau} - \arg(s_t^* \mathbf{v}_{t-1}^H \mathbf{y}_t))}{a_t + \rho|s_t|^2}}{(\pi^{t+1} (a_t + \rho|s_t|^2))^M}.$$

By further marginalization of ϕ_t and with the definitions

$$\begin{aligned} A_1 &\triangleq \frac{e^{-\|\mathbf{x}\|^2 - \sum_{\tau=1}^t \|\mathbf{y}_\tau\|^2 + \frac{\rho(\|\mathbf{v}_{t-1}\|^2 + |s_t|^2 \|\mathbf{y}_t\|^2)}{a_t + \rho|s_t|^2}}}{(\pi^{t+1} (a_t + \rho|s_t|^2))^M}, \\ A_2 &\triangleq \frac{2\rho|s_t^* \mathbf{v}_{t-1}^H \mathbf{y}_t|}{a_t + \rho|s_t|^2}, \\ A_3 &\triangleq \sum_{\tau=1}^{t-1} \phi_\tau - \arg(s_t^* \mathbf{v}_{t-1}^H \mathbf{y}_t), \end{aligned}$$

the likelihood $p(\mathbf{x}, \mathbf{y}_1^t | \phi_1^{t-1}, \phi_t, s_1^{t-1}, s_t)$ is given by

$$p(\mathbf{x}, \mathbf{y}_1^t | \phi_1^{t-1}, \phi_t, s_1^{t-1}, s_t) = A_1 \left(\alpha_0 I_0(A_2) + 2 \sum_{p=1}^{\infty} \alpha_p I_p(A_2) \cos(pA_3) \right). \quad (76)$$

Finally the detection rule is

$$\begin{aligned} \hat{s}_t = \arg \max_{s_t \in \mathcal{S}} & \frac{\rho(\|\mathbf{v}_{t-1}\|^2 + |s_t|^2 \|\mathbf{y}_t\|^2)}{a_t + \rho|s_t|^2} - M \ln(a_t + \rho|s_t|^2) \\ & + \ln \left(\alpha_0 I_0 \left(\frac{2\rho|\chi[t]|}{a_t + \rho|s_t|^2} \right) + 2 \sum_{p=1}^{\infty} \alpha_p I_p \left(\frac{2\rho|\chi[t]|}{a_t + \rho|s_t|^2} \right) \cos \left(p \left(\sum_{\tau=1}^{t-1} \phi_\tau - \arg(\chi[t]) \right) \right) \right), \end{aligned} \quad (77)$$

where $\chi[t] \triangleq s_t^* \mathbf{v}_{t-1}^H \mathbf{y}_t$.

8.5 Numerical Examples

In this section, we provide some numerical examples to investigate the performance of the various derived detectors. In all the examples, the phase noise increments $\phi_m[t]$, $m = 1, \dots, M$, $t = 1, \dots, T$ are assumed to be i.i.d. zero mean wrapped Gaussian random variables with variance σ_ϕ^2 . Hence,

the pdf of the increments is given by (58) with $\alpha_{m,p} = \exp\left(-\frac{\sigma_\phi^2}{2}\right)^{p^2}$. Also in all the Figures $M = 20$ and the length of the data interval is $T = 20$. In Fig. 9 the performance of the genie-aided receiver for the synchronous operation in (68) is compared with the suboptimal decision-feedback detector for the non-synchronous operation in (61) for the constant channel case.

The input symbols are selected from a QPSK constellation with equal probability and the variance of the phase noise increments is $\sigma_\phi^2 = 0.1$. In Fig. 10 the uncoded SER performance of the detectors (68) and (61) is shown as a function of ρ for equiprobable 8-PSK symbols. In Figs. 11 and 12 the uncoded SER performance of the detectors in (71) and (77) is plotted as a function of ρ . In Fig. 11 the symbols are QPSK with equal probability and $\sigma_\phi^2 = 0.07$, and in Fig. 12 the symbols are equiprobable 8-PSK symbols with $\sigma_\phi^2 = 0.01$. In all cases shown in Figs. 9-12, it is clear that the decision-feedback non-synchronous detector outperforms the corresponding genie-aided synchronous detector. This establishes the validity of the results to setups with more data channel uses.

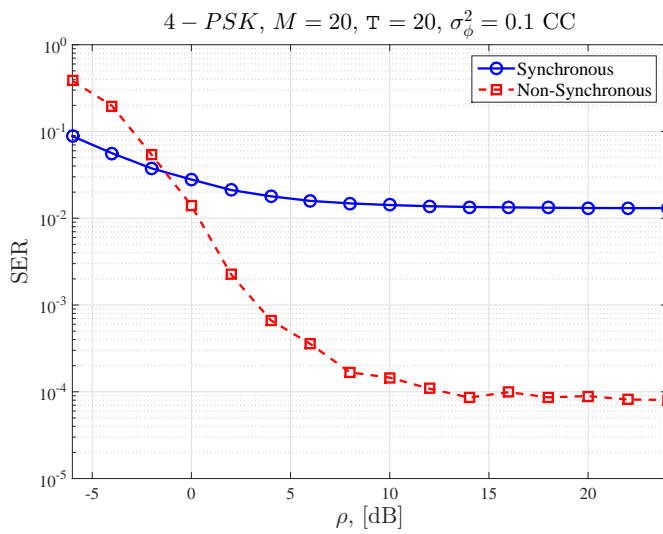


Figure 9: Uncoded SER vs ρ for QPSK symbols, $T = 20$, $M = 20$ and $\sigma_\phi^2 = 0.1$ for the constant channel.

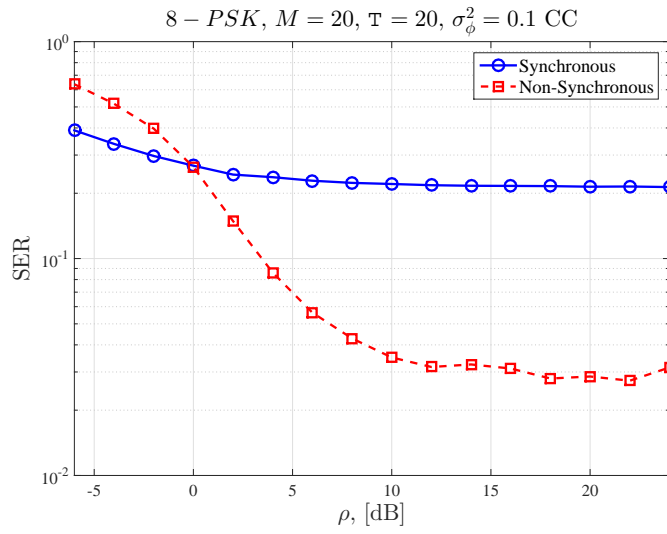


Figure 10: Uncoded SER vs ρ for 8-PSK symbols, $T = 20$, $M = 20$ and $\sigma_\phi^2 = 0.1$ for the constant channel.

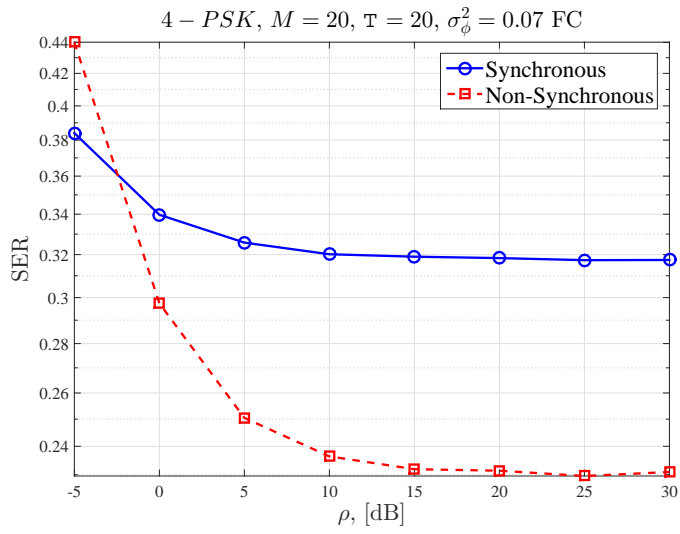


Figure 11: Uncoded SER vs ρ for QPSK symbols, $T = 20$, $M = 20$ and $\sigma_\phi^2 = 0.07$ for the fading channel.

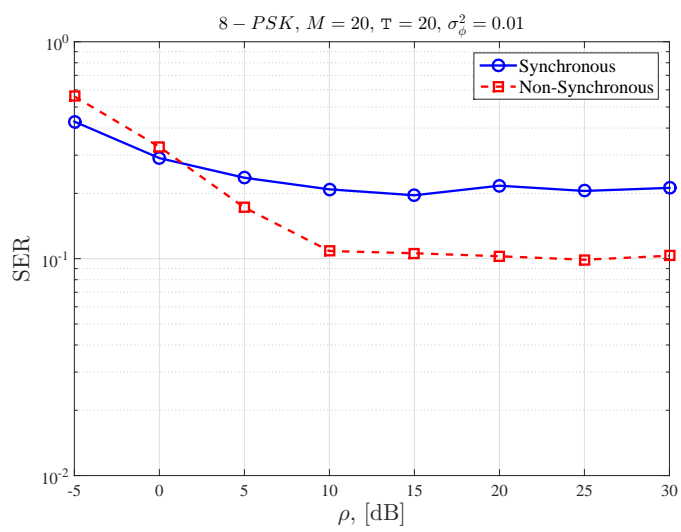


Figure 12: Uncoded SER vs ρ for 8-PSK symbols, $T = 20$, $M = 20$ and $\sigma_\phi^2 = 0.01$ for the fading channel.

Bibliography

- [1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, pp. 3590–3600, Nov. 2010.
- [2] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, pp. 40–60, Jan. 2013.
- [3] E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, pp. 186–195, Feb. 2014.
- [4] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Transactions on Communications*, vol. 61, pp. 1436–1449, Apr. 2013.
- [5] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?," *IEEE Journal on Selected Areas in Communications*, vol. 31, pp. 160–171, Feb. 2013.
- [6] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Transactions on Information Theory*, vol. 60, pp. 7112–7139, Nov. 2014.
- [7] E. Björnson, M. Matthaiou, and M. Debbah, "Massive MIMO with non-ideal arbitrary arrays: Hardware scaling laws and circuit-aware design," *IEEE Transactions on Wireless Communications*, vol. 14, pp. 4353–4368, Aug. 2015.
- [8] A. Lapidath, "On phase noise channels at high SNR," in *Proc. of the 2002 IEEE Information Theory Workshop*, pp. 1–4, Oct. 2002.

- [9] G. Durisi, "On the capacity of the block-memoryless phase-noise channel," *IEEE Communications Letters*, vol. 16, pp. 1157–1160, Aug. 2012.
- [10] G. Durisi, A. Tarable, C. Camarda, R. Devassy, and G. Montorsi, "Capacity bounds for MIMO microwave backhaul links affected by phase noise," *IEEE Transactions on Communications*, vol. 62, pp. 920–929, Mar. 2014.
- [11] M. Khanzadi, G. Durisi, and T. Eriksson, "Capacity of SIMO and MISO phase-noise channels with common/separate oscillators," *IEEE Transactions on Communications*, vol. 63, pp. 3218–3231, Sept. 2015.
- [12] A. Pitarokoilis, S. K. Mohammed, and E. G. Larsson, "Uplink performance of time-reversal MRC in massive MIMO systems subject to phase noise," *IEEE Transactions on Wireless Communications*, vol. 14, pp. 711–723, Feb. 2015.
- [13] A. Pitarokoilis, S. K. Mohammed, and E. G. Larsson, "Achievable rates of ZF receivers in massive MIMO with phase noise impairments," in *Proc. of the 47th Asilomar Conference on Signals, Systems and Computers, 2013*, pp. 1004–1008, Nov. 2013.
- [14] A. Viterbi, "Optimum detection and signal selection for partially coherent binary communication," *IEEE Transactions on Information Theory*, vol. 11, pp. 239–246, Apr. 1965.
- [15] G. J. Foschini, R. D. Gitlin, and S. B. Weinstein, "On the selection of a two-dimensional signal constellation in the presence of phase jitter and Gaussian noise," *Bell System Technical Journal*, vol. 52, pp. 927–965, July-Aug. 1973.
- [16] P. Y. Kam, S. S. Ng, and T. S. Ng, "Optimum symbol-by-symbol detection of uncoded digital data over the Gaussian channel with unknown carrier phase," *IEEE Transactions on Communications*, vol. 42, pp. 2543–2552, Aug. 1994.
- [17] M. Khanzadi, D. Kuylenstierna, A. Panahi, T. Eriksson, and H. Zirath, "Calculation of the performance of communication systems from measured oscillator phase noise," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 61, pp. 1553–1565, May 2014.
- [18] G. Colavolpe, A. Barbieri, and G. Caire, "Algorithms for iterative decoding in the presence of strong phase noise," *IEEE Journal on Selected Areas in Communications*, vol. 23, pp. 1748–1757, Sept. 2005.

- [19] R. Krishnan, G. Colavolpe, A. Graell i Amat, and T. Eriksson, "Algorithms for joint phase estimation and decoding for MIMO systems in the presence of phase noise and quasi-static fading channels," *IEEE Transactions on Signal Processing*, vol. 63, pp. 3360–3375, July 2015.
- [20] R. Krishnan, M. Khanzadi, T. Eriksson, and T. Svensson, "Soft metrics and their performance analysis for optimal data detection in the presence of strong oscillator phase noise," *IEEE Transactions on Communications*, vol. 61, pp. 2385–2395, June 2013.
- [21] H. Mehrpouyan, A. A. Nasir, S. D. Blostein, T. Eriksson, G. K. Karagiannidis, and T. Svensson, "Joint estimation of channel and oscillator phase noise in MIMO systems," *IEEE Transactions on Signal Processing*, vol. 60, pp. 4790–4807, Sept. 2012.
- [22] T. Höhne and V. Ranki, "Phase noise in beamforming," *IEEE Transactions on Wireless Communications*, vol. 9, pp. 3682–3689, Dec. 2010.
- [23] R. Krishnan, M. Khanzadi, N. Krishnan, Y. Wu, A. Graell i Amat, T. Eriksson, and R. Schober, "Linear Massive MIMO precoders in the presence of phase noise – a large-scale analysis," *IEEE Transactions on Vehicular Technology*, vol. PP, no. 99, pp. 1–1, 2015.
- [24] E. Björnson, M. Matthaiou, A. Pitarokoilis, and E. G. Larsson, "Distributed Massive MIMO in cellular networks: Impact of imperfect hardware and number of oscillators," in *23rd European Signal Processing Conference (EUSIPCO), 2015*, pp. 2436–2440, Aug. 2015.
- [25] P. Hou, B. Belzer, and T. Fischer, "Shaping gain of the partially coherent additive white gaussian noise channel," *IEEE Communications Letters*, vol. 6, pp. 175–177, May 2002.
- [26] A. Pitarokoilis, E. Björnson, and E. G. Larsson, "Optimal detection in training assisted SIMO systems with phase noise impairments," in *IEEE International Conference on Communications (ICC), 2015*, pp. 2597–2602, June 2015.
- [27] K. Mardia and P. Jupp, *Directional Statistics*. Wiley Series in Probability and Statistics, Wiley, 2009.
- [28] S. R. Jammalamadaka and A. Sengupta, *Topics in Circular Statistics*. World Scientific Pub Co Inc, 2001.

- [29] B. Goebel, R.-J. Essiambre, G. Kramer, P. J. Winzer, and N. Hanik, "Calculation of mutual information for partially coherent Gaussian channels with applications to fiber optics," *IEEE Transactions on Information Theory*, vol. 57, pp. 5720–5736, Sept. 2011.
- [30] A. Demir, A. Mehrotra, and J. Roychowdhury, "Phase noise in oscillators: a unifying theory and numerical methods for characterization," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 47, pp. 655–674, May 2000.
- [31] H. L. van Trees, *Detection, Estimation, and Modulation Theory, Part I*. New York: Wiley Interscience, 2001.
- [32] A. Laforgia, "Bounds for modified Bessel functions," *Journal of Computational and Applied Mathematics*, vol. 34, pp. 263–267, Apr. 1991.
- [33] D. J. Bordelon, "Problem 72-15," *SIAM Review*, vol. 15, pp. 665–670, Jul. 1973.
- [34] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*. Elsevier/Academic Press, Amsterdam, 7th ed., 2007.
- [35] B. Hochwald, T. Marzetta, and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Transactions on Information Theory*, vol. 50, pp. 1893–1909, Sept. 2004.
- [36] G. Bennett, "Probability inequalities for the sum of independent random variables," *Journal of the American Statistical Association*, vol. 57, pp. 33–45, Mar. 1962.
- [37] P. Billingsley, *Probability and Measure, Anniversary Edition*. Wiley, 2012.
- [38] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. Cambridge, UK: Cambridge Univ. Press, 2004.
- [39] A. Viterbi, *Principles of Coherent Communication*. McGraw-Hill Inc., 1966.

PAPER E

Performance of the Massive MIMO Uplink with OFDM and Phase Noise

Refereed article submitted to the IEEE Communications Letters

©2016 IEEE. The layout has been revised.

Performance of the Massive MIMO Uplink with OFDM and Phase Noise

Antonios Pitarokoilis, Emil Björnson, and Erik G. Larsson

Abstract

The performance of multi-user Massive MIMO-OFDM uplink systems in the presence of base station (BS) phase noise impairments is investigated. Closed-form achievable rate expressions are rigorously derived under two different operations, namely the case of a common oscillator (synchronous operation) at the BS and the case of independent oscillators at each BS antenna (non-synchronous operation). It is observed that the non-synchronous operation exhibits superior performance due to the averaging of intercarrier interference. Further, radiated power scaling laws are derived, which are identical to the phase-noise-free case.

1 Introduction

The demand for more data traffic in cellular systems has been the driving force for research in the field during the last decades. Base station (BS) densification and allocation of more bandwidth have been pivotal to handling more traffic, but this is expected to change in the future. Already, BSs with multiple antennas have been deployed to increase the performance in terms of energy and spectral efficiency. Recently, *Massive multiple-input multiple-output (MIMO)* systems [1], i.e., systems where the BS is equipped with an unprecedentedly large number of antennas and communicates with a few tens of non-cooperative users over the same time and frequency resources, has become one of the most prominent candidates for the evolution of cellular systems.

The realization of affordable Massive MIMO requires the use of a large number of inexpensive, and potentially low-quality, components. Fortunately, it has been shown that Massive MIMO systems are, in general, robust against hardware impairments [2]. One of the most important hardware impairments is phase noise (PN), which is a random and varying phase rotation of the information signal. PN arises in coherent communication systems due to the various noise sources in the circuits of the local oscillators (LOs) that are used for the modulation of the information signal from the baseband to passband at the transmitter and vice versa at the receiver.

The effect of PN has been extensively studied under various systems models. The effect of PN in Massive MIMO systems has also been recently investigated for single-carrier systems [3,4]. However, most contemporary wireless systems employ orthogonal frequency division multiplexing (OFDM), i.e., they are multi-carrier. PN destroys the orthogonality of subcarriers in OFDM and, therefore, degrades significantly the system performance [5]. PN-impaired Massive MIMO-OFDM systems appear to be less well understood. To the authors' knowledge, the only available study of the problem is [6]. In [6], rate expressions for a single-user Massive MIMO-OFDM uplink were given that are valid under specific assumptions and in the asymptotic limit when the number of base station antennas goes to infinity. In the present work, we revisit the PN-impaired Massive MIMO-OFDM uplink, but for a multi-user system with autonomous, spatially multiplexed terminals, and develop rigorous lower bounds on ergodic capacity that are valid for any finite number of antennas.

The contributions of this work can be summarized as follows. First, the frequency-domain system model definition is based on the time-domain description of the channel, which gives a parsimonious description of the problem under study. Further, achievable rates are rigorously derived under the assumptions of perfect CSI at the BS and MRC processing. Closed-form expressions for the proposed achievable rates are provided, which hold for any finite number of BS antennas. Two distinct LO operations are considered, i.e., the synchronous operation and the non-synchronous operation, and compared. Based on the derived closed-form expressions, radiated power scaling laws are derived and compared with previous studies on single-carrier systems.

2 System Model

A BS equipped with M antenna elements communicates in the uplink with K non-cooperative single-antenna users. The complex baseband equivalent channel impulse response (CIR) of each user to the BS is assumed to be frequency-selective and is modeled as a finite impulse response (FIR) filter with L sample-spaced taps, where T_s is the sampling period. The l -th channel gain of the k -th user to the m -th antenna element is noted as $g_{m,k,l} = \sqrt{d_{k,l}}h_{m,k,l}$. The coefficients $h_{m,k,l} \sim \mathcal{CN}(0,1)$ are i.i.d. circularly symmetric complex Gaussian random variables and correspond to the small scale fading. The sequence of positive reals $d_{k,0}, \dots, d_{k,L-1}$ corresponds to the power delay profile (PDP) of the channel from user k to the BS antenna array and it is the same to all the BS antenna elements. The PDPs are assumed to be constant and known and the following condition holds

$$\mathbb{E} \left[\left| \sum_{l=0}^{L-1} g_{m,k,l} \right|^2 \right] = \sum_{l=0}^{L-1} d_{k,l} = \beta_k. \quad (1)$$

The fast fading coefficients are assumed to remain constant for a time interval T_{coh} and then they change to an independent realization. OFDM transmission is used, where N_c is the number of subcarriers, N_{CP} is the length of the cyclic prefix (CP) in samples. Hence, an OFDM symbol is $N_{\text{OFDM}} = N_c + N_{\text{CP}}$ samples long. It is assumed that n_{coh} OFDM symbols span the interval $T_{\text{coh}} = n_{\text{coh}}N_{\text{OFDM}}T_s$.

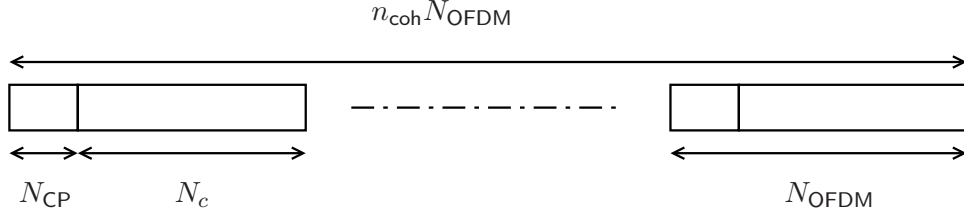


Figure 1: Structure of a coherence interval in the time domain.

The received signal is impaired by PN at the receiver, $\phi_m[i]$. Free-running LOs are assumed, hence the PN realizations are generated by Wiener processes, i.e., the PN realization at the m -th BS antenna at time i is

$$\phi_m[i] = \phi_m[i-1] + \sigma_\phi w_m[i], \quad (2)$$

where $w_m[i] \sim \mathcal{N}_{\mathbb{R}}(0, 1)$ is a sequence of independent and identically distributed real Gaussian random variables. The constant σ_ϕ^2 is the variance of the increments of the PN processes, which is a measure of the quality of the LOs ($\sigma_\phi^2 = 0$ for ideal LOs). Two distinct operations are considered, namely, the case where the PN processes are independent among the BS antennas (non-synchronous operation) and the case where the same PN process impairs all the BS antennas (synchronous operation), i.e., $w_1[i] \equiv \dots \equiv w_M[i]$. The non-synchronous operation corresponds to a distributed deployment, where each BS antenna element is equipped with a different LO and the synchronous operation corresponds to a centralized scenario, where the same LO is used for all the BS antennas.

The received signal at time i at the m -th BS antenna, when the time domain information symbol sequences $x_k[i-L+1], \dots, x_k[i]$ are sent, is given by

$$y_m[i] = \sqrt{\rho} \sum_{k=1}^K \sum_{l=0}^{L-1} \sqrt{d_{k,l}} h_{m,k,l} e^{-j\phi_m[i]} x_k[i-l] + z_m[i], \quad (3)$$

where $z_m[i] \sim \mathcal{CN}(0, 1)$ additive white Gaussian noise (AWGN). The time-domain samples received at the BS within the n -th OFDM symbol, $1 \leq n \leq n_{\text{coh}}$, after the CP is removed are given by

$$\mathbf{y}(n) = \sqrt{\rho} \sum_{k=1}^K \Phi(n) \mathbf{G}_k \mathbf{x}_k(n) + \mathbf{z}(n), \quad (4)$$

where $\mathbf{y}(n) \triangleq [\mathbf{y}[n, 1]^T, \dots, \mathbf{y}[n, N_c]^T]^T \in \mathbb{C}^{MN_c}$,

$$\mathbf{y}[n, i] \triangleq \begin{bmatrix} y_1[nN_{\text{OFDM}} - N_c + i] \\ \vdots \\ y_M[nN_{\text{OFDM}} - N_c + i] \end{bmatrix} \in \mathbb{C}^M \text{ and}$$

$$\mathbf{x}_k(n) \triangleq [x_k[nN_{\text{OFDM}} - N_c + 1], \dots, x_k[nN_{\text{OFDM}}]]^T \in \mathbb{C}^{N_c}.$$

The matrix \mathbf{G}_k is block circulant with each block being a column vector of length M , corresponding to the time-domain CIRs of the user k to the BS. In particular, the (ν_1, ν_2) M -column vector of \mathbf{G}_k is given by

$$[\mathbf{G}_k]_{\nu_1, \nu_2} = \begin{cases} \mathbf{g}_{k, (\nu_1 - \nu_2) \bmod N_c}, & \text{if } 0 \leq (\nu_1 - \nu_2) \bmod N_c \leq L - 1 \\ \mathbf{0}, & \text{otherwise.} \end{cases}$$

Finally, $\Phi(n) = \text{blkdiag}\{\Phi[n, 1], \dots, \Phi[n, N_c]\}$ and $\Phi[n, i] = \text{diag}\{e^{-j\phi_1[nN_{\text{OFDM}} - N_c + i]}, \dots, e^{-j\phi_M[nN_{\text{OFDM}} - N_c + i]}\}$. The frequency-domain received signal is given by

$$\mathbf{y}(n) = \sqrt{\rho} \sum_{k=1}^K \mathbf{G}_k(n) \mathbf{x}_k(n) + \mathbf{z}(n), \quad (5)$$

where $\mathbf{y}(n) \triangleq (\mathbf{F}_{N_c} \otimes \mathbf{I}_M) \mathbf{y}(n)$, $\mathbf{x}_k(n) \triangleq \mathbf{F}_{N_c} \mathbf{x}_k(n)$, $\mathbf{z}(n) \triangleq (\mathbf{F}_{N_c} \otimes \mathbf{I}_M) \mathbf{z}(n)$ are the frequency-domain counterparts of the corresponding time-domain quantities and $\mathbf{G}_k(n) \triangleq (\mathbf{F}_{N_c} \otimes \mathbf{I}_M) \Phi(n) \mathbf{G}_k \mathbf{F}_{N_c}^H$ is the effective frequency-domain channel matrix for user k . By \mathbf{F}_{N_c} we denote the unitary DFT matrix of size N_c and \otimes is the Kronecker product.

3 Achievable Rates

The physical propagation channel from user k , which we denote¹ as $\mathbf{G}_k = (\mathbf{F}_{N_c} \otimes \mathbf{I}_M) \mathbf{G}_k \mathbf{F}_{N_c}^H = \text{blkdiag}\{\mathbf{g}_{k,1}, \dots, \mathbf{g}_{k,N_c}\}$, is assumed to be known perfectly at the BS. The vector $\mathbf{g}_{k,p} \in \mathbb{C}^{M \times 1}$ is the frequency response of

¹We stress the notational difference between the frequency-domain physical propagation channel, \mathbf{G}_k , and the effective frequency-domain channel matrix with PN, $\mathbf{G}_k(n)$, at the n -th OFDM symbol for user k .

the propagation channel of user k at the p -th subcarrier. With maximum ratio combining (MRC) the detected information vector for user k , $\hat{\mathbf{x}}_k(n)$ is given by

$$\hat{\mathbf{x}}_k(n) = \mathbf{G}_k^H \mathbf{y}(n) = \sqrt{\rho} \mathcal{G}_k(n) \mathbf{x}_k(n) + \text{EN}_k(n), \quad (6)$$

where $\mathcal{G}_k(n) \triangleq \mathbb{E} [\mathbf{G}_k^H \mathbf{G}_k(n)]$ and

$$\begin{aligned} \text{EN}_k(n) &= \sqrt{\rho} (\mathbf{G}_k^H \mathbf{G}_k(n) - \mathcal{G}_k(n)) \mathbf{x}_k(n) \\ &\quad + \sqrt{\rho} \sum_{\substack{p=1 \\ p \neq k}}^K \mathbf{G}_k^H \mathbf{G}_p(n) \mathbf{x}_p(n) + \mathbf{G}_k^H \mathbf{z}(n). \end{aligned} \quad (7)$$

An achievable rate, $R_k(n)$, for the k -th user at the n -th OFDM symbol is given by Proposition 1.

Proposition 1. *With $\mathbf{x}_k(n) \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$, an achievable rate for the k -th user at the n -th OFDM symbol with MRC processing when the propagation channel \mathbf{G}_k is available at the receiver is given by*

$$R_k(n) = \log_2 |\mathbf{I}_{N_c} + \rho \mathcal{G}_k^H(n) \mathbf{C}_k^{-1}(n) \mathcal{G}_k(n)|, \quad (8)$$

using the methodology in [7], where $\mathbf{C}_k(n)$ is the covariance matrix of the effective noise term $\text{EN}_k(n)$.

Remark 1. *We note that the derived achievable rate is a function of the OFDM symbol index, n . Observe that the effective channel matrix, $\mathcal{G}_k(n)$, and the effective noise statistics are also functions of n . Hence, the rate in Proposition 1 for the n -th OFDM symbol can be achieved by using a dedicated, for the n -th OFDM symbol, Gaussian codebook with codewords that span over multiple coherence intervals of the channel fading. Similar coding strategies have earlier been proposed in [3].*

Definition 1. *The effective, over the coherence interval, net rate per subcarrier for the k -th user, \hat{R}_k , is defined as*

$$\hat{R}_k \triangleq \frac{1}{n_{\text{coh}}} \sum_{n=1}^{n_{\text{coh}}} \hat{R}_k(n) = \frac{1}{n_{\text{coh}}} \sum_{n=1}^{n_{\text{coh}}} \frac{1}{N_c} \frac{N_c}{N_{\text{OFDM}}} R_k(n), \quad (9)$$

where $\hat{R}_k(n)$ is the effective net rate per subcarrier at the n -th OFDM symbol. The fraction $\frac{N_c}{N_{\text{OFDM}}}$ accounts for the rate loss due to the CP and $\frac{1}{N_c}$ is the per-subcarrier normalization.

In the following we provide closed-form expressions for the achievable rates in Proposition 1 for the two different operations: synchronous and non-synchronous.

Proposition 2. For both operations the effective channel matrix $\mathbf{G}_k(n)$ is given by $\mathbf{G}_k(n) = MN_c \mathbf{\Psi}(n)$, where

$$\mathbf{\Psi}(n) \triangleq (\mathbf{F}_{N_c} \bar{\mathbf{\Phi}}(n) \mathbf{F}_{N_c}^H) \circ (\mathbf{F}_{N_c} \mathbf{D}_k \mathbf{F}_{N_c}^H),$$

$$\bar{\mathbf{\Phi}}(n) \triangleq \text{diag} \left\{ e^{-\frac{\sigma_\phi^2}{2} |nN_{\text{OFDM}} - N_c + 1|}, \dots, e^{-\frac{\sigma_\phi^2}{2} |nN_{\text{OFDM}}|} \right\},$$

$\mathbf{D}_k = \text{diag} \{d_{k,0}, \dots, d_{k,L-1}, \mathbf{0}\}$ and \circ is the element-wise Hadamard product.

Proposition 3. The covariance matrix of the effective noise is given by

$$\mathbf{C}_k^\times(n) = \rho M \mathbf{C}_{1,k} + \rho M(M-1) \mathbf{C}_{2,k}^\times(n) - \rho \mathbf{G}_k(n) \mathbf{G}_k^H(n) + M \beta_k \left(\rho \sum_{q \neq k}^K \beta_q + 1 \right) \mathbf{I}_{N_c}, \quad (10)$$

where $\times = s$ for synchronous operation and $\times = ns$ for non-synchronous operation,

$$\begin{aligned} [\mathbf{C}_{1,k}]_{\nu_1, \nu_2} &= \mathbf{w}_{\nu_1}^H \left(\mathbf{D}_{k, \nu_1}^H \mathbf{D}_{k, \nu_2} \otimes \tilde{\mathbf{\Phi}} \right) \mathbf{w}_{\nu_2} + \frac{\beta_k \mathbf{d}_{\nu_2 - \nu_1}^*}{N_c} \mathbf{w}_{\nu_1}^H \left(\mathbf{I}_{N_c} \otimes \tilde{\mathbf{\Phi}} \right) \mathbf{w}_{\nu_2}, \\ [\mathbf{C}_{2,k}^s(n)]_{\nu_1, \nu_2} &= \mathbf{w}_{\nu_1}^H \left(\mathbf{D}_{k, \nu_1}^H \mathbf{D}_{k, \nu_2} \otimes \tilde{\mathbf{\Phi}} \right) \mathbf{w}_{\nu_2}, \\ [\mathbf{C}_{2,k}^{ns}(n)]_{\nu_1, \nu_2} &= N_c^2 [\mathbf{F}_{N_c} \bar{\mathbf{\Phi}}(n) \mathbf{F}_{N_c}^H \mathbf{D}_{k, \nu_1}^H \mathbf{D}_{k, \nu_2} \mathbf{F}_{N_c} \bar{\mathbf{\Phi}}(n) \mathbf{F}_{N_c}^H]_{\nu_1, \nu_2}, \\ \mathbf{w}_{\nu_3, \nu_2} &\triangleq \left[1 \ e^{-j \frac{2\pi(\nu_3 - \nu_2)}{N_c}} \ \dots \ e^{-j \frac{2\pi(\nu_3 - \nu_2)(N_c - 1)}{N_c}} \right]^T, \ \mathbf{w}_{\nu_2} \triangleq \left[\mathbf{w}_{1, \nu_2}^T \ \dots \ \mathbf{w}_{N_c, \nu_2}^T \right]^T, \\ [\tilde{\mathbf{\Phi}}]_{\nu_4, \nu_5} &= e^{-\frac{\sigma_\phi^2}{2} |\nu_4 - \nu_5|}, \ \mathbf{D}_{k, \nu_1} \triangleq \text{diag} \{d_{k, 1 - \nu_1}, \dots, d_{k, N_c - \nu_1}\} \text{ and } \mathbf{d}_{k, \nu_2 - \nu_1} \triangleq \\ &[\mathbf{F}_{N_c} \mathbf{D}_k \mathbf{F}_{N_c}^H]_{\nu_1, \nu_2}. \end{aligned}$$

Remark 2. The matrix $\mathbf{C}_2^\times(n)$ corresponds to part of the interference caused by the uncertainty of the effective channel gain, $\mathbf{G}_k(n)$, due to PN. The elements of the matrix $\mathbf{C}_2^s(n)$ are not a function of n , however, the elements of the matrix $\mathbf{C}_2^{ns}(n)$ have magnitudes which decay with n and N_{OFDM} as $e^{-\sigma_\phi^2 n N_{\text{OFDM}}}$. This implies that part of the interference in the non-synchronous operation averages out, whereas the same part in the synchronous operation does not. Hence, the achievable rates are higher for the non-synchronous operation than the synchronous. Similar observations were drawn in other studies on single-carrier systems [3, 4].

Corollary 1 can be derived from Propositions 2 and 3 in the absence of PN.

Corollary 1. *In the absence of PN the effective channel matrix is given by $\mathbf{G}_k(n) = M\beta_k\mathbf{I}_{N_c}$ and the covariance of the effective noise is $\mathbf{C}_k = M\beta_k\left(\rho\sum_{q=1}^K\beta_q + 1\right)\mathbf{I}_{N_c}$. Consequently, the rate for the k -th user at the n -th OFDM symbol is given by*

$$R_k^{np}(n) \geq N_c \log_2 \left(1 + \frac{\rho M \beta_k}{\rho \sum_{q=1}^K \beta_q + 1} \right), \quad \forall n. \quad (11)$$

Proposition 4. *For a fixed desired achievable rate, it is possible to reduce the total radiated power by 3 dB for every doubling of the number of BS antennas, M , i.e., an $O(M)$ radiated power gain is achievable in the presence of PN.*

Proof. Let $\rho = E_u/M$. Then, as $M \rightarrow \infty$, the rate converges to the limiting value

$$R_k^{\times,\text{lim}}(n) = \log \left| \mathbf{I}_{N_c} + E_u \mathbf{\Psi}^H(n) \left(\mathbf{C}_k^{\times,\text{lim}}(n) \right)^{-1} \mathbf{\Psi}(n) \right|, \quad (12)$$

where $\mathbf{C}_k^{\times,\text{lim}}(n) \triangleq E_u \left(\mathbf{C}_k^{\times}(n) - N_c^2 \mathbf{\Psi}(n) \mathbf{\Psi}^H(n) \right) + \beta_k \mathbf{I}_{N_c}$. \square

The scaling law given by Proposition 4 is the same as the one that holds in the PN-free case.

4 Numerical Examples

In this section we present some numerical examples that highlight the degradation introduced by PN at Massive MIMO-OFDM systems. In all the examples the following parameters are held constant $N_c = 128$, $N_{\text{CP}} = 16$, $N_{\text{OFDM}} = 144$, $n_{\text{coh}} = 5$, $L = 17$, $K = 10$. The PDPs of all users are assumed to be exponential with $d_{k,l} = \beta_k e^{-\alpha_k l} / \zeta_k$, where $\boldsymbol{\beta} = [\beta_1, \dots, \beta_K]^T = [0.9940, 0.5852, 0.6289, 0.6984, 0.5370, 0.8420, 0.7012, 0.9914, 0.7011, 0.8103]^T$, $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_K]^T = [1.0557, 0.6844, 0.7120, 0.5773, 1.4138, 1.2067, 1.0578, 0.8134, 0.6662, 1.1225]^T$ and ζ_k is a normalization factor such that (1) holds. The vectors $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ were generated from uniform distributions $\mathcal{U}[0.5, 1]$ and $\mathcal{U}[0.5, 1.5]$ and kept fixed for the generation of the figures.

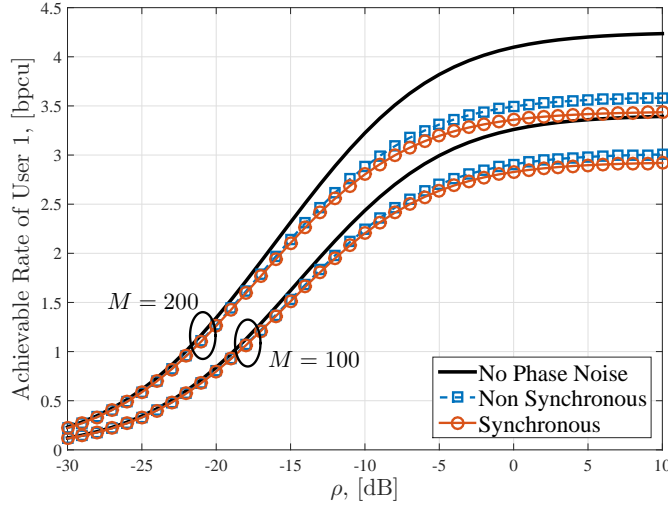


Figure 2: Rate of user 1 as a function of ρ for $M = [100, 200]$ and $\sigma_\phi^2 = 10^{-4}$.

In Fig. 2 the effective net rate per subcarrier of user 1, \hat{R}_k , given in Definition 1, is shown for $M = [100, 200]$ and $\sigma_\phi^2 = 10^{-4}$. It is clear that the achievable rates increase with M and the performance of non-synchronous operation is superior to the synchronous. This can be attributed to the fact that the part of the interference, $\mathbf{C}_2^{ns}(n)$, due to the uncertainty of the effective channel matrix, $\mathbf{G}_k(n)$, in the non-synchronous operation averages out, as noted in Remark 2. The PN-free bound from (11) is also plotted as a benchmark.

In Fig. 3 the effective net rate per subcarrier of user 1, $\hat{R}_k(n)$, is plotted for the first and the last OFDM symbol of the coherence interval for $\sigma_\phi^2 = 10^{-3}$. The progressive degradation of the performance is attributed to the partial loss of coherency between the physical channel, \mathbf{G}_k , and the effective channel, $\mathbf{G}_k(n)$, due to the evolution of the PN processes. It is further observed that, for the particular choice of parameters, the performance difference is more pronounced at the first OFDM symbol rather than the last. In the initial part of the coherence interval and the effective channel has not drifted significantly from the propagation channel and the effective channel gain is relatively high. In addition, the averaging behavior of a part of the interference in the non-synchronous operation is significant already in the first OFDM symbol. Hence, the difference in achievable rate performance between the synchronous and non-synchronous operation is substantial. On

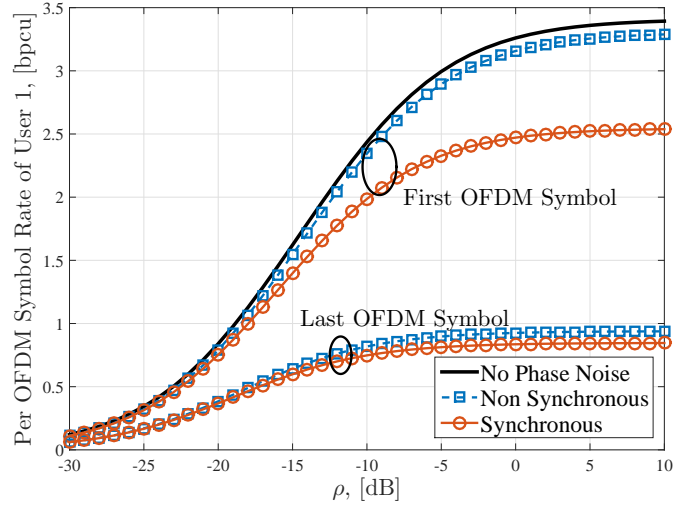


Figure 3: Progressive degradation of achievable rate for the first and n_{coh} -th OFDM symbol for $\sigma_{\phi}^2 = 10^{-3}$.

the other hand, at the last OFDM symbol the drift due to PN is significant and the effective channel gain is greatly reduced for both operations. In this case the averaging of part of the interference in the non-synchronous operation does not yield the same gains in performance as in the first OFDM symbol.

In Fig. 4 the effective net rate per subcarrier of user 1 is plotted as a function of M . However, in this case $\rho = E_u/M$, where $E_u = 0$ dB is constant. The rate in all cases saturates at a non-zero limiting value. This establishes the claim in Proposition 4.

5 Conclusion

The effect of PN in Massive MIMO-OFDM uplink channels is studied and closed-form achievable rates are rigorously derived for two different operations: synchronous and non-synchronous LOs in the array. It is shown that the non-synchronous operation is superior due to the averaging of the interference due to effective channel uncertainty. A progressive degradation of the achievable rate performance is observed and a radiated power

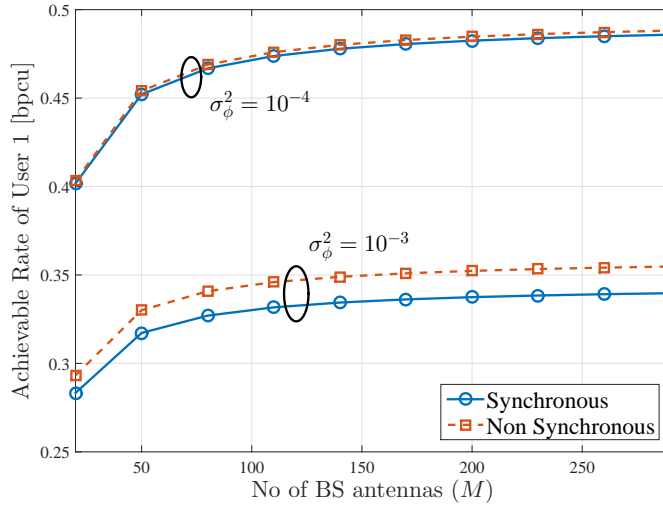


Figure 4: Radiated power scaling law as given by Proposition 4.

scaling law is provided. The behavior observed here resembles the performance observed in earlier single-carrier studies [3, 4]. The in-depth study of the same system model under PN at the users and with estimated CIRs is of particular interest. However, it increases substantially the complexity and the notation of the problem and, therefore, will be part of the future work based on this initial study.

Bibliography

- [1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 3590–3600, Nov. 2010.
- [2] E. Björnson *et al.*, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Trans. Inf. Theory*, vol. 60, pp. 7112–7139, Nov. 2014.
- [3] A. Pitarokoilis, S. K. Mohammed, and E. G. Larsson, "Uplink performance of time-reversal MRC in massive MIMO systems subject to phase noise," *IEEE Trans. Wireless Commun.*, vol. 14, pp. 711–723, Feb. 2015.
- [4] E. Björnson, M. Matthaiou, and M. Debbah, "Massive MIMO with non-ideal arbitrary arrays: Hardware scaling laws and circuit-aware design," *IEEE Trans. Wireless Commun.*, vol. 14, pp. 4353–4368, Aug. 2015.
- [5] T. Pollet, M. Van Bladel, and M. Moeneclaey, "BER sensitivity of OFDM systems to carrier frequency offset and wiener phase noise," *IEEE Trans. Commun.*, vol. 43, pp. 191–193, Feb/Mar/Apr 1995.
- [6] R. Krishnan *et al.*, "On the impact of oscillator phase noise on the uplink performance in a massive MIMO-OFDM system," *CoRR*, vol. abs/1405.0669, 2014.
- [7] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inf. Theory*, vol. 49, pp. 951 – 963, Apr. 2003.

Linköping Studies in Science and Technology
Dissertations, Division of Communication Systems
Department of Electrical Engineering (ISY)
Linköping University, Sweden

Erik Axell, *Spectrum Sensing Algorithms Based on Second-Order Statistics*, Dissertation No. 1457, 2012.

Tumula V. K. Chaitanya, *HARQ Systems: Resource Allocation, Feedback Error Protection, and Bits-to-Symbol Mappings*, Dissertation No. 1526, 2013.

Johannes Lindblom, *The MISO Interference Channel as a Model for Non-Orthogonal Spectrum Sharing*, Dissertation No. 1555, 2013.

Reza Moosavi, *Improving the Efficiency of Control Signaling in Wireless Multiple Access Systems*, Dissertation No. 1556, 2014.

Mirsad Čirkić, *Efficient MIMO Detection Methods*, Dissertation No. 1570, 2014.

Hien Quoc Ngo, *Massive MIMO: Fundamentals and System Designs*, Dissertation No. 1642, 2015.

Anu Kalidas Muralidharan Pillai, *Signal Reconstruction Algorithms for Time-Interleaved ADCs*, Dissertation No. 1672, 2015.