

Adaptive Modulation Technique for Broadband Communication in Indoor Optical Wireless Systems

vorgelegt von
Diplom-Ingenieurin
Jelena Vučić geb. Grubor
aus Belgrad, Serbien

von der Fakultät IV - Elektrotechnik und Informatik
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Ingenieurwissenschaften
- Dr.-Ing. -

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender:	Prof. Dr.-Ing. Adam Wolisz
1. Gutachter:	Prof. Dr.-Ing. Klaus Petermann
2. Gutachter:	Prof. Dr.-Ing. Andreas Kirstädter

Tag der wissenschaftlichen Aussprache: 09. Juli 2009

Berlin 2009

D 83

Acknowledgement

Thank you, mama i tata, for your endless love and unconditional support.

Thank you, Nikola, for your love and tolerance, but above all thank you for your patience.

Thank you, Klaus, for watching my back and not letting me give up. Thank you for all your guidance “on and off the court”.

Thank you, Volker, for your support and advice throughout my PhD years. Thank you for your generosity.

Thank you, Lutz, thank you, Kai, for your all-around expertise which saved my day on innumerable occasions.

Thank you, Christoph, for performing measurements and for being a great help in the OMEGA project.

Thank you, Professor Petermann, for your advice on my research, and especially for valuable comments on my dissertation.

Thank you, Professor Kirstädter, for agreeing to officially review the thesis.

Thank you, Luz, Michael, Matthias, Lutz, Kai, and Markus, for your great spirits and big hearts.

Thank you, colleagues, for being more than that. You made me feel welcome from day one.

Thank you, friends, for being just that.

To you all I owe my deepest gratitude.

Abstract

Optical wireless (OW) is an attractive and viable technology for high-speed short-range (indoor) communication, with a good prospect of taking part in the wider fourth generation (4G) vision. This thesis investigates and assesses the potentials of the *adaptive modulation technique* in an OW orthogonal frequency division multiplex (OFDM) based system for a typical indoor wireless local area network (WLAN) scenario. Even though the regarded technique is known from digital subscriber line (DSL) systems and is well researched for radio systems, specific properties of the optical wireless channel call for a separate evaluation of its performance in OW links.

The adopted channel model reveals presence of large dynamics in channel bandwidth and magnitude, which indicates that it is difficult to maintain the signal-to-noise ratio sufficient for high bit rates under all channel conditions using reasonable transmitter power levels. An adaptive OFDM system, which is designed to dynamically adjust the transmission rate to the current channel state, by varying the order of modulation and power of the individual subcarriers is, therefore, an attractive solution to efficiently exploit the channel capacity, while maintaining a reliable link.

Deployment of adaptive OFDM in an OW system brings about various additional benefits. System complexity is relocated from optical front-ends to electrical signal processing. In this way, large spatial coverage can be provided even with very simple and inexpensive transceivers, without the need for pointing and tracking mechanisms. Within a cell, user mobility is supported and the transmission link is made robust to shadowing, while reliable connectivity is maintained thanks to dynamic rate adaptation. The possibility of complete digitalization makes signal processing very efficient, so that it can be implemented on a logic device and integrated on a microchip. OFDM inherently deals with inter-symbol-interference (major degradation factor of the high-speed OW links) and allows for the simplest linear equalization at the receiver.

In this thesis, questions related to the system capacity, as the theoretical limit for rate performance, as well as achievable rates in practical systems, are in focus. The deployment of loading algorithms which apply different orders of square quadrature-amplitude-modulation formats while complying with a certain error rate, is investigated. As a method to improve the system power-efficiency and further enhance the transmission rate on account of an acceptable increase in error rate, the influence of controlled clipping on system performance is examined. Main issues regarding the opto-electrical front-ends for the considered communication scenario are discussed, in order to arrive at a simple design, which at the same time allows broadband transmission with sufficient link budget.

Kurzfassung

Optische Freistrahübertragung ist eine attraktive und nützliche Technologie für hochratige Inhaus-Kommunikation (über kürzeren Distanzen), mit guten Aussichten zum Einsatz in der vierten Generation der Mobilkommunikationssysteme. In der vorliegenden Dissertation wird das Leistungsvermögen der adaptiven Modulationstechnik in einem auf orthogonalem Frequenzmultiplex (OFDM) basierenden optischen Freistrah-System für typische Inhaus-WLAN-Anwendungen untersucht und bewertet. Obwohl diese Technik schon für DSL- (Digital Subscriber Line) und Funkkanal-Anwendungen bekannt ist, erfordern die besonderen Bedingungen in einem optischen Freistrah-Kanal eine eigenständige Untersuchung.

Das zu Grunde liegende Kanalmodell zeigt eine hohe Dynamik in der Bandbreite und Amplitude der Kanalantwort. Dies hat zur Folge, dass es anspruchsvoll ist, für hohe Übertragungsraten ein ausreichend hohes Signal-zu-Rausch-Verhältnis zu gewährleisten - und zwar für alle vorkommenden Zustände des Kanals und für realistische Pegel der Sendeleistung. Ein adaptives OFDM-System ist so aufgebaut, dass die Übertragungsrate durch Verändern der Modulationsordnungen und Signalleistungen der individuellen Subträger an den momentanen Kanalzustand angepasst werden kann. Auf diese Weise wird die Kanalkapazität effizient genutzt und gleichzeitig die Beibehaltung einer robusten Verbindung sichergestellt.

Der Einsatz des adaptiven OFDM-Verfahrens in einem optischen Freistrah-System bringt viele zusätzliche Vorteile mit sich. Die Haupt-Komplexität des Systems wird konzentriert in der elektrischen Signalverarbeitung und ermöglicht den Einsatz einfacher und preiswerter optischer Frontends. Dadurch kann man eine große Abdeckung des Raumes erreichen, ohne dabei z.B. zusätzliche Mechanismen zur Ausrichtung und Nachführung der Transceiver verwenden zu müssen. Zudem wird die Mobilität des Nutzers innerhalb einer Kommunikations-Zelle unterstützt und die Übertragung robust gemacht gegen Abschattung (Unterbrechung des direkten Übertragungspfades), wobei eine betriebssichere Verbindung erhalten bleibt - dank der dynamischen Adaption der Übertragungsrate. Des Weiteren ermöglicht eine komplett digitalisierte elektronische Signalverarbeitung, die schaltungstechnisch in einen Mikrochip integriert werden kann, ein sehr effizientes elektrisches Transceiver-Design. Das OFDM-Verfahren bekämpft inhärent die Inter-Symbol-Interferenz (ein wesentlicher Störfaktor in hochratigen optischen Freistrah-Verbindungen), und erlaubt die einfache lineare Entzerrung am Empfänger.

Die vorliegende Arbeit behandelt insbesondere Fragen bezüglich der Systemkapazität (theoretische Grenzen der Übertragungsrate) und der praktisch erreichbaren Übertragungsraten. Dabei wird der Einsatz von Zuweisungsalgorithmen untersucht, die bei einer vorgegebenen Bitfehlerrate des Systems individuellen Subträgern verschiedene QAM-Modulationsordnungen und Signalleistungen zuordnen.

Als Methode zur Verbesserung der System-Leistungseffizienz und zur weiteren Erhöhung der Übertragungsrate wird der Einsatz des kontrollierten Signal-Clippings vor der optischen Quelle untersucht. Wesentliche Fragestellungen in Bezug auf die Gestaltung der opto-elektrischen Frontends für das vorgesehene Anwendungsszenario werden ebenfalls diskutiert. Das Ziel ist hier, zu einem einfachen Design zu kommen, welches zugleich eine breitbandige Übertragung bei ausreichendem Link-Budget unterstützt.

Contents

Mathematical Symbols and Acronyms	III
1 Introduction	1
1.1 Basic link designs and state-of-the-art	4
1.2 Goal and contributions of the thesis	7
1.3 Outline of the thesis	9
2 Communication Scenario and Channel	11
2.1 Indoor communication scenario	11
2.2 Propagation effects in the OW channel	12
2.3 Noiseless OW channel model	15
2.3.1 LOS component	16
2.3.2 Diffuse component	19
2.3.3 Composite (LOS plus diffuse) channel	22
2.4 Noise sources and mechanisms in the OW channel	27
3 OFDM-Based Rate-Adaptive OW System	31
3.1 Mathematical modeling of noiseless multipath channels	31
3.2 Multicarrier transmission and OFDM	33
3.3 OFDM transmission chain	37
3.4 Rate-adaptive OW system	42
4 Design of Simple Opto-Electrical Front-Ends	49
4.1 Optical source	49
4.1.1 Safety considerations	50
4.1.2 Transmission of analogue multitone signals	55
4.2 Infrared receiver	66
4.2.1 Optical concentrator and optical filter	67
4.2.2 Photodetector and preamplifier	72
4.2.3 Summary of chosen parameters	86
5 Evaluation of System Capacity	89
5.1 Definitions and constraints	90

5.1.1	Capacity of the narrowband power- and band-limited Gaussian channel	90
5.1.2	Capacity of the wideband (frequency selective) power- and band-limited Gaussian channel	91
5.1.3	Constraints on the transmit signal in OW systems	93
5.2	TMS method for MSM-based OW systems	94
5.2.1	Trigonometric moment space method	94
5.2.2	Derivation of the capacity upper bound	97
5.2.3	Conservative, straight-forward constraint	106
5.2.4	Geometric interpretation on a 2D example	110
5.3	Performance of water-filling in MSM-based OW systems	112
6	Throughput Maximization in Practical Systems	119
6.1	Rate enhancement using loading algorithms	119
6.1.1	Bit- and power-loading concept	120
6.1.2	Krongold algorithm for power-limited systems	123
6.1.3	Conservative approach to loading in an OW system	125
6.1.4	Upper-bounding the conservative loading solution	128
6.2	Controlled clipping	130
6.2.1	Influence on error performance	134
6.2.2	Rate maximization	140
6.2.3	Capacity estimate	147
6.2.4	Implementation of power control	150
6.2.5	Summary of main results	152
6.3	System coherence interval	153
6.3.1	Variation in the OW channel	155
6.3.2	Quantification of channel and SNR deviation	159
6.3.3	Consequences on link design	163
7	Summary and Outlook	167
A	Parameters for Ray Tracing Simulations	177
B	Elimination of ISI with OFDM	179
C	Radiation Safety Considerations	183
C.1	Maximum permissible Tx power based on AELs for eye safety . . .	183
C.2	Maximum permissible Tx power based on MPEs for skin safety . .	188
D	Non-Imaging Concentrator Gain	189
	Bibliography	190

Mathematical Symbols and Acronyms

Mathematical Symbols

A_{bud}	Amplitude budget, sum of symbol constellation amplitudes of the independent subcarriers (positive frequencies only)
A_{eff}	Effective receiver surface
A_{eq}	Sum of mean constellation amplitudes on the positive frequencies after the power distributing algorithm
A_n	Amplitude of n^{th} subcarrier symbol constellation
$\langle A_n \rangle$	Mean amplitude of n^{th} subcarrier symbol constellation
$A_{n,\text{max}}$	Maximum amplitude of n^{th} subcarrier symbol constellation
A_{Room}	Surface of the model room
$A_{\text{Rx}}, A_{\text{PD}}$	Receiver surface (bare photodetector)
a	Aperture diameter of the optical source
a_n	Real part of the n^{th} subcarrier symbol constellation
$a_{n,c}$	Real part of the clipping error on the n^{th} subcarrier
B	Signal bandwidth in baseband, Effective noise bandwidth
B_{C}	Channel coherence bandwidth
B_{D}	Doppler spread
B_{S}	Signal bandwidth
B_{SC}	Bandwidth of one subchannel, subcarrier
BER_{CL}	Overall bit-error-rate resulting from clipping
BER_{LA}	Overall bit-error-rate (Gaussian noise)
BER_{TOT}	Overall bit-error-rate (noise and clipping)
b_n	Imaginary part of the n^{th} subcarrier symbol constellation
$b_{n,c}$	Imaginary part of the clipping error on the n^{th} subcarrier
C	Channel (system) capacity in [bit/s]
C_{D}	Discrete channel (system) capacity in [bit/channel use]
C_{det}	Photodetector capacitance (shunt capacitance)
C_{FET}	FET gate capacitance
C_{Kron}	Upper bound on the rates achievable by the conservative Krongold algorithm

C_{stray}	Stray capacitance
C_{TOT}	Total detector capacitance (photodiode plus preamplifier)
\mathcal{CH}_k	k-dimensional space where the complex sequences fulfil the conditions of the optical wireless channel
\mathcal{CPT}_k	k-dimensional cross-polytope
c	Speed of light in vacuum $c = 3 \cdot 10^8$ m/s
$c(k)$	Clipped portion of the transmit signal
$\text{conv}(\cdot)$	Convex hull of a space curve
D	Channel delay spread
d	Photodetector diameter
$\det(\cdot)$	Matrix determinant
$\mathbb{E}\{\cdot\}$	Mathematical expectation
$E_n(x - x_0)$	Error of the channel coefficient corresponding to the n^{th} subcarrier between Rx positions x and x_0
$F(M)$	Excess noise factor
FOV_{det}	Field-of-view of a (bare) detector
f	Modulating frequency
f_g	Total photodetector bandwidth
f_M	Multiplication bandwidth
$f_{\text{max}}(\text{MF}_n)$	Ratio between the n^{th} subcarrier symbol constellation mean electrical power and square of the maximum amplitude
f_n	n^{th} subcarrier frequency, n^{th} fundamental tone
f_{RC}	RC -circuit bandwidth
f_{source}	Source frequency
f_{tr}	Transit time bandwidth
f_0	3 dB cut-off frequency
$f_{3\text{dB}}$	3 dB cut-off frequency of the composite channel
$f_{\diamond}(\text{MF}_n)$	Ratio between the n^{th} subcarrier symbol constellation mean electrical power and square of the mean amplitude
G	Gain of the inverting amplifier
$[\mathbf{G}_{\text{CP}}]$	Cyclic prefix generation matrix
\mathcal{GE}	Generalized ellipse space curve
$g(P_n)$	Constraint function
$g(\psi)$	Gain of the optical concentrator
g_m	FET transconductance
g_{max}	Maximum gain of the optical concentrator
$H(f)$	Channel frequency response
$[\mathbf{H}_{\text{eq}}]$	Matrix of equivalent channel frequency response
$H_{\text{LD-PD}}$	Attenuation between LD input and PD output
$H_n, H(f_n)$	OW channel frequency response coefficient at f_n
h, l, w	Height, length and width of the model room
h'	Minimum Tx-Rx distance for the considered scenario
$[\mathbf{h}]$	Matrix of (discrete) channel impulse response

$h(k)$	Discrete channel impulse response (also denoted by $h_D(k)$)
$h(t)$	Channel impulse response
$h^2(\xi)$	Optical wireless channel delay power spectrum
$[\mathbf{h}_{\text{pc}}]$	Pseudocirculant channel matrix
$[\mathbf{h}_0], [\mathbf{h}_1]$	Parts of $[\mathbf{h}]$ (see $[\mathbf{h}]$)
$\langle I \rangle$	Electrical current of the laser operating point
$I(t), i(t)$	Electrical current - large signal, small signal
$I(\phi)$	Radiant intensity in direction ϕ
$\Im\{\cdot\}$	Imaginary part of the argument
$[\mathbf{I}_N]$	Identity matrix
I_{pc}	Average photocurrent
I_{th}	Laser threshold current
I_0	Maximum radiant intensity (in direction $\phi = 0$)
$\langle i^2 \rangle$	Mean power of small signal, signal power
K	K -factor (ratio between LOS and diffuse signals)
K_{LD}	Number of elements in a laser array
k_A	Ionization coefficient ratio
k_B	Boltzmann constant $k_B = 1.38 \cdot 10^{-23}$ Ws/K
L	Cyclic prefix length (in samples)
$L(P_n)$	Lagrange function
L_C	Channel length (in samples)
M	Multiplication factor (average)
M	Order of the QAM modulation format
\mathcal{M}_k	k -dimensional trigonometric moment space
M_{opt}	Optimal multiplication factor
M_0	Multiplication factor at low frequencies
m	Lambert radiation index
m_{eff}	Effective modulation index
m_{max}	Maximum modulation index
m_n	Modulation index of the n^{th} tone (subcarrier)
N	Number of independent subcarriers (including DC)
N_{CNT}	Number of OFDM symbols for control information
N_D	Number of OFDM symbols for part of frame carrying data
N_F	Number of fundamental tones at the optical source input
N_{FB}	Number of OFDM symbols for feed-back link
N_n	Effective noise power in subchannel n
N_{SC}	I- and Q- noise components variance in one subchannel
N_T	Number of OFDM symbols for training sequence
N_{used}	Number of active subcarriers in the system
N_0	AWGN single-sided power spectral density
n	Refractive index of the optical concentrator
\mathcal{O}	Order of the output signal component
P	OFDM symbol length (in samples), block length

$\langle P \rangle$	Optical power of the laser operating point
P_{bud}	Budget of modulating power, sum of symbol powers of the independent subcarriers (positive frequencies)
P_{E}	Electrical power (variance) of the transmitted signal
$P_{\text{el,in}}$	Electrical power of the tone at the optical source input
$P_{\text{el,out}}$	Electrical power of the tone after detection
P_{eq}	Signal modulating power (over positive frequencies) after the amplitude distributing algorithm
P_n	n^{th} subcarrier power (electrical) of the symbol constellation
P_{O}	Limit on the mean optical power at the transmitter
$P_{\text{ref,bud}}$	Reference power budget for the power control mechanism
P_{Rx}	Optical power at the receiver
P_{TOT}	Total electrical power corresponding to R_{TOT} (see R_{TOT})
P_{Tx}	Optical power at the transmitter
p	Additional margin for noise
p_{BG}	Bright-sky-light spectral irradiance of $p_{\text{BG}} = 5.8 \mu\text{W}/\text{cm}^2\text{nm}$
q	Electricron charge $q = 1.6 \cdot 10^{-19} [\text{As}]$
\mathbb{R}^k	k-dimensional real-valued space
R_b	Bit rate
R_{C}	Radius of the hemispherical concentrator
$[\mathbf{R}_{\text{CP}}]$	Cyclic prefix removal matrix
$\Re\{\cdot\}$	Real part of the argument
R_{FB}	Feedback resistor in transimpedance preamplifier
R_{in}	Input resistance of the preamplifier circuit
R_{L}	Load resistor in low impedance preamplifier
R_n	Spectral efficiency of the n^{th} subcarrier
R_{sam}	Sampling rate
R_{sym}	Symbol rate
R_{TOT}	Total spectral efficiency resulting from the LA
r	Distance between Tx and Rx
r_{max}	Maximum Tx-Rx distance for the considered scenario
r_{PD}	Radius of the circle-shaped photodetector
S	Optical source opto-electrical conversion factor
\mathcal{S}_k	k-dimensional symplex
\mathcal{SQ}_k	k-dimensional space where the complex sequences fulfil the conservative conditions of the optical wireless channel
SER_{CL}	Overall bit-error-rate resulting from clipping
SER_{LA}	Overall bit-error-rate resulting from Gaussian noise
SER_{TOT}	Overall bit-error-rate (both Gaussian noise and clipping)
S_2	Second laser characteristic derivative in the operating point
S_3	Third laser characteristic derivative in the operating point
T	Temperature
$T(x)$	Diffuse signal delay (to LOS) at Rx (x is Rx position)

$T(\psi)$	Transmission coefficient of the optical filter
T_C	Channel coherence time
T_{CNT}	Control information duration within transmission frame
T_{CP}	Length (duration) of cyclic prefix
T_D	Duration of the data block in one transmission frame
T_{FB}	Feed-back link duration within transmission frame
T_{FFT}	OFDM symbol length w/o CP
$T_{\text{LOS}}(x)$	LOS signal propagation delay with Rx at position x
$T_{\text{LOS},0}$	LOS signal propagation delay with Rx directly below Tx
T_{OFDM}	OFDM symbol length with CP
T_{sam}	Sampling interval
T_{sym}	Signal period, Symbol interval, OFDM symbol w/o CP
T_T	Duration of the training sequence in one transmission frame
T_0	Maximum transmission coefficient of the optical filter
t	Time
$\langle t \rangle$	Average time between two reflections
$u(t)$	Heaviside step function
V	Space volume
V_{Room}	Volume of the model room
v	Relative speed of the objects
v_d	Minority carrier saturation drift velocity (illuminated layer)
$[\mathbf{W}_N]$	Vandermonde matrix (for FFT)
w	Thickness of the i-zone (depletion zone)
$[\mathbf{X}]$	Input vector of sent symbols (f -domain)
$[\hat{\mathbf{X}}]$	Estimated vector after equalization at the Rx (f -domain)
$X_n, X(n)$	Complex symbol sent on the n^{th} subcarrier (f -domain)
X_0	Input at IFFT on zero frequency
$[\mathbf{x}]$	Signal vector at the channel input (t -domain)
$x(k)$	k^{th} sample of signal at the channel input (t -domain)
$x_c(k)$	Transmit signal samples after clipping
$\langle x^2 \rangle$	See $\langle i^2 \rangle$
$[\mathbf{Y}]$	Output vector of received symbols (f -domain)
$Y_n, Y(n)$	Complex symbol received on the n^{th} subcarrier (f -domain)
$[\mathbf{y}]$	Signal vector at the channel output (t -domain)
$y(k)$	k^{th} sample of signal at the channel output (t -domain)
Z_{LD}	Differential impedance of the optical source
Z_n	Spectral component of complex AWGN on n^{th} subcarrier
Z_{PD}	Serial resistance of the photodiode
$[\mathbf{z}]$	Vector of noise samples at the channel output (t -domain)
α	Lagrange multiplier
α_e	Impact ionization coefficient of electrons
α_h	Impact ionization coefficient of holes
$\beta_{i,k}^{\text{CE}}$	Channel coefficient variation between Rx positions i and k

$\beta_{i,k}^{\text{LA}}$	SNR variation between Rx positions i and k
$\beta_k^{\text{CE(LA)}}$	Channel (SNR) variation when Rx moves k positions away
β_n	Optimal slope parameter for the n^{th} subcarrier on OPC
Γ	FET channel noise factor
γ_g	Geometric mean of γ -factors of the active subcarriers
γ_n	γ -factor of the n^{th} subcarrier (analogous to SNR)
Δf	Subcarrier spacing
ΔT	Delay between LOS and diffuse signal contributions
ΔT_g	Global delay between LOS and diffuse signals when Rx is moved from the center to the corner of the model room
$\Delta T_{\text{LOS,g}}$	Global delay between LOS signal arrivals when receiver is moved from the center to the corner of the model room
$\Delta \lambda$	Optical filter bandwidth
$\Delta \phi_n$	LOS signal phase variation at frequency f_n ch6)
$\delta(t)$	Dirac function
ε_r	Relative permittivity
ε_0	Vacuum permittivity $\varepsilon_0 = 8.85 \cdot 10^{-12} \text{Fm}^{-1}$
η	Photodetector responsivity
η_{DIFF}	Diffuse signal path gain (loss)
η_{LOS}	LOS signal path gain (loss)
$\eta_{\text{LOS},0}$	LOS signal amplitude with Rx directly below Tx
θ	Angle of incidence upon the optical filter
θ_C	Field of view of the optical filter (angular bandwidth)
θ_{max}	Maximum angle of incidence on the optical filter
θ_t	Truncation angle of the optical concentrator
λ	Wavelength (signal wavelength)
μ	Mean channel delay
ξ	Delay
ρ	Clipping probability
$\langle \rho \rangle$	Average reflectivity of all surfaces in the model room
ρ_{sim}	Clipping probability obtained directly by simulations
σ^2	Total noise variance (power)
σ_{FET}^2	FET noise variance (power)
σ_{S}^2	Shot noise variance (power)
σ_{T}^2	Thermal noise variance (power)
τ^{-1}	Decay constant of the exponential function
τ_{M}	Effective transit time in the avalanche region
τ_{tr}	Carrier transit time in the i-zone
$\Phi(\cdot)$	cdf of a normalized normal distribution
ϕ	Angle between Tx orientation and LOS direction
ϕ_n	Phase of the symbol constellation sent on the n^{th} subcarrier
ϕ_{max}	Maximum radiation angle for the considered scenario
$\phi_{1/2}$	Radiation half-angle

Ψ_C	Field of view of the optical concentrator
ψ	Angle of incidence at the detector
Ω	Solid angle
ω_0	Base angular frequency
$(\cdot)!$	Factorial
$\lfloor \cdot \rfloor$	First integer smaller than the argument
$[\cdot]^T$	Matrix transpose operation
$[\cdot]^*$	Matrix conjugation operation

Acronyms

AC	Alternating Current
ADC	Analogue-To-Digital Converter
AEL	Accessible Emission Limit
APD	Avalanche Photo-Diode
AR	Anti-Reflection
AWGN	Additive White Gaussian Noise
BER	Bit-Error Rate
BPSK	Binary Phase Shift Keying
C/I	Carrier-To-Interference Ratio
CP	Cyclic Prefix
CPT	Cross-Polytope
CSI	Channel State Information
CW	Continuous Wave
DAC	Digital-To-Analogue Converter
DC	Direct Current
DFE	Decision Feedback Equalizer
DFT	Discrete Fourier Transform
DIFF	Diffuse
DSL	Digital Subscriber Line
DSP	Digital Signal Processing
FDM	Frequency Division Multiplex
FEC	Forward Error Correction
FET	Field-Effect Transistor
FFT	Fast Fourier Transform
FOV	Field-Of-View
FPGA	Field Programmable Gate Array
HD	Harmonic Distortion
I/Q	In-Phase/Quadrature
IDFT	Inverse Discrete Fourier Transform
IFFT	Inverse Fast Fourier Transform
IM/DD	Intensity Modulation Direct Detection
IMP	Inter-Modulation Product

IR	Infrared
IrDA	Infrared Data Association
ISI	Inter-Symbol-Interference
LA	Loading Algorithm
LD	Laser Diode
LED	Light-Emitting Diode
LOS	Line Of Sight
LPF	Low-Pass Filter
LTI	Linear Time-Invariant
LUT	Look-Up-Table
MC	Monte Carlo
MCM	Multicarrier Modulation
MF	Modulation Format
MPE	Maximal Permissible Exposure
MSM	Multiple-Subcarrier Modulation
NEV	Noise Enhancement Vector
O/E	Opto-Electrical Conversion
OFDM	Orthogonal Frequency Division Multiplex
OOK	On-Off Keying
OPC	Operational Characteristic
OW	Optical Wireless
P/S	Parallel-To-Serial Conversion
PD	Photo-Diode
PIN	P-doped-Intrinsic-N-doped
PPM	Pulse-Position Modulation
QAM	Quadrature-Amplitude Modulation
RF	Radio Frequency
Rx	Receiver
S/P	Serial-To-Parallel Conversion
SCI	System Coherence Interval
SER	Symbol-Error Rate
Si	Silicon
SIK	Sequence Inversion Keying
SIR	Signal-To-Interference Ratio
SISO	Single Input Single Output
SNCR	Signal-To-Noise-And-Clipping-Interference Ratio
SNR	Signal-To-Noise Ratio
TIA	Trans-Impedance Amplifier
TMS	Trigonometric Moment Space
Tx	Transmitter
UWB	Ultra Wideband
WLAN	Wireless Local Area Network
WPAN	Wireless Personal Area Network

Chapter 1

Introduction

Together with the demand for bandwidth provision required by advanced applications, the end-users are growing accustomed to comfortability of wireless access. In addition, with wireless connections, network providers can be more relaxed with respect to the network planning, installation, maintenance, expansion, etc. Even though behind the term wireless, radio frequencies (RF) are conventionally considered, both industry and scientific community agree that the fourth generation (4G) of wireless communication systems will be based on several complementary access technologies. The main reason is that each of these technologies is tailored to a particular typical access scenario, and hence using just one technology would hardly be optimal in all cases.

Alternatively to radio, wireless transmission can be realized with lightwaves (visible or infrared) propagating through the air. For outdoor applications, this technology is referred to as *free space optics*, FSO [1], whereas the term *optical wireless* (OW) *communication* is customary for indoor links. Infrared (IR) links can use several transmission windows in the near-IR range (750-1600 nm) which exhibit favorable attenuation characteristics. However, the window around 850 nm wavelength and the window around 1550 nm are usually regarded as the most attractive ones, due to the availability of inexpensive components and robustness with respect to eye-safety, respectively. Since recently, there is also an increasing amount of research in the area of *visible light communication*, VLC [2–4]. Here, light emitting diodes (LEDs), whose primary functions are illumination and signalling, are simultaneously used for information transmission. The focus of this thesis is on infrared radiation based OW communication.

Optical wireless communication offers some attractive features, distinct from those of radio. Lightwaves occupy frequencies of hundreds of THz (radio frequencies are up to 100 GHz), and therefore theoretically have the potential for very high data rates. At the same time, the radio spectrum is getting increasingly crowded and electromagnetic interference (EMI) becomes a serious problem. As a conse-

quence, different bodies regulate the spectrum world-wide, so that among other rules, license fees are mandatory. Due to the limited resources, dedicated channels are usually narrow-band and broadband transmission for densely populated buildings is challenging. Originally, for Wireless Local Area Network (WLAN), robust spread spectrum techniques were deployed, which provided a maximum data rate of 11 Mbit/s (IEEE Standard 802.11b [5]). Nowadays, IEEE 802.11g WLAN standard foresees additionally an orthogonal frequency division multiplex (OFDM) based physical layer, with a maximum bit-rate of 54 Mbit/s [5]. The release of 802.11n amendment is awaited which will increase the rates up to 300 Mbit/s thanks to the multiple-input-multiple-output (MIMO) technology [6]. The second standard worth mentioning here is Bluetooth, whose current version 2.0 offers 3 Mbit/s maximum rates for short-range applications, based on spread spectrum technology [7]. In order to cope with all the demand and insufficiency of the available bandwidth, the Ultra-Wideband (UWB), as well as 60 GHz radio technologies are considered. In UWB systems, the information is spread over a large bandwidth (~ 500 MHz) and transmitted with a very low power level, which in theory remains below the noise level of the conventional radio receivers. In that way, the spectrum can be shared with other users [8]. Due to the huge bandwidth requirement of UWB links, high data rates are possible only over short distances (few meters). Nevertheless, out of concern that the rise in noise level by a number of UWB transmitters would endanger the stability of the existing communication services, many are reluctant to take on a risk with this technology. In the context of standardization, UWB has been proposed for Wireless Personal Area Networks (WPAN) and appeared in the IEEE 802.15.3a PAN standard draft. However, after several years of deadlock between two technology approaches (multi-band OFDM and spread spectrum), the task group has been dissolved [9]. At the millimeter-wave frequency bands (60 GHz), RF circuit design and implementation seem to be quite complex. Currently, a task group is working towards a proposal of a millimetre-wave-based alternative physical layer for the IEEE high-rate WPAN standard [10].

Quite the opposite, the enormous bandwidth of optical links is still available and unregulated (free). Moreover, there is no electromagnetic interference with systems operating in the radio spectrum. Another property of wireless optical links with several beneficial consequences, is that lightwaves do not penetrate through opaque surfaces. In indoor environments, this inherently defines a room as a communication cell. Since networks with opaque cell borders don't suffer from inter-cell interference, the same operating optical frequencies can be used in every cell, leading to high aggregate network capacity. Moreover, since transmissions in different rooms do not have to be coordinated, all units in the cellular architecture can be identical, which greatly simplifies the network design. Another advantage over radio links is the enhanced privacy and better security against eavesdropping.

Whereas radio may be a favored media for applications where large coverage and high mobility are aimed, deployment of OW technology is interesting for high-speed short-range (indoor) links. It is particularly attractive for environments where radio systems are not desired, where interference-free radio spectrum is hardly available, where wireless transmission with enhanced security degree is required (banks, military objects, hospitals, airports, . . .), or when international compatibility is targeted. If used complementary in the above mentioned heterogeneous wireless access concept, the two media provide a broad range of applications that one technology alone would find difficult to do.

Several performance degrading effects need to be considered in an OW link. The first effect is *optical path loss*. It presents the part of the transmitted power, which is lost or not captured by the receiver aperture. In indoor links, however, there is no environmental degradation (due to fog, particulate matter, . . .), which is a significant issue in free-space optics.

Further on, as in radio systems, between the source and the receiver in a confined space (room), two distinct modes of wave propagation are possible - via line-of-sight (LOS) and via reflections. Differently to radio waves, lightwaves reflect diffusely at most surfaces (walls and objects) in the room, rendering a multitude (continuum) of possible transmission paths. The effect of *multipath propagation* on the OW signal is similar to that on the radio signal - the received electric field experiences amplitude fades (on a scale of wavelength) and the signal is dispersed. However, for OW communication, intensity modulation with direct detection (IM/DD) is considered as the only practicable transmission method.¹ Due to square-law detection on a detector whose relative size is huge with respect to the incoming signal wavelength, the effects of multipath fading are negligible in OW, which greatly simplifies the link design. Multipath dispersion, however, presents a major degradation factor at higher transmission bandwidths (rates). This channel effect is modeled by a linear time-invariant baseband impulse response.

Background light, coming from the sun and/or artificial lighting (fluorescent and incandescent lamps), is usually the dominant source of noise in an OW link. It primarily induces shot-noise in the photodetector which can be modeled as additive white gaussian noise (AWGN) due to its high intensity, but fluorescent lamps also cause interference on frequencies up to several tens of kHz. When the background light level is low, receiver preamplifier noise becomes dominant.

¹It seems extremely difficult to collect appreciable signal power in a single electromagnetic mode. Thus, a spatially incoherent reception makes it difficult to construct an efficient heterodyne or homodyne receiver. Since polarization and direction of the incoming mode must be actively tracked, the high complexity of coherent receivers is not favorable in a low-cost system (even though they could render a much better sensitivity). Coherent receivers for OW links were considered in, e.g., [11] or [12].

Even though the model of an OW link, characterized with multipath propagation, mathematically looks the same as for radio or wired electrical channels (linear system), there are several important qualitative differences. In particular, because of intensity modulation, the electrical signal is directly converted into optical power and transmitted through the air. Since optical power cannot be negative, the signal in front of the optical source should take only non-negative values, as well. Moreover, because of the unguided transmission, it is required to take special care of the safety of radiation levels. Regulations pose a constraint on the mean optical power at the transmitter, which limits the first moment (i.e., mean value) of the transmitted signal, rather than its second moment (i.e., power), as is common for wired or radio channels.

1.1 Basic link designs and state-of-the-art

According to the mentioned propagation modes, two generic types of OW links are defined. As illustrated in Fig. 1.1 a, a *directed* or LOS link relies on the existence of an uninterrupted light path between a highly directed transmitter (Tx) and a narrow field-of-view (FOV) receiver (Rx). In a *diffuse* link (Fig. 1.1 b), the connection between a wide-beam Tx and a large-field-of-view Rx relies on numerous signal reflections off the surfaces in the room, instead.

LOS links experience minimal path loss, they are relatively free from multipath dispersion and are able to diminish the influence of ambient light. Commercially, such links are present for many years in low bit-rate, simplex remote control applications for domestic electrical equipment (TV, stereo,...), and are standardized by the Infrared Data Association (IrDA) [13]. Today, IrDA interfaces can be found also in a variety of portable devices, like mobile phones, laptop computers and personal digital assistants. In its most recent version (fast IrDA), the standard offers rates up to 16 Mbit/s (over approximately 1 m range), with even higher data rates foreseen for the future (100 Mbit/s with ultra fast IrDA). If LOS is uninterrupted, link performance is limited only by the available power budget and hence very high transmission rates (beyond 100 Mbit/s) are shown to be possible. On the other hand, LOS links require careful alignment of transceivers and generically provide very small coverage.

In order to enhance the coverage and support the user mobility, different pointing and tracking mechanisms were proposed in combination with arrays of Tx-Rx modules. In [14], 1 Gbit/s link was demonstrated with mechanical steerable optics, and in [14–17], rates up to 155 Mbit/s per terminal were achieved with pointing and tracking mechanisms based on electronics and supported by an imaging infrared receiver. However, such mechanisms are complex, bulky and costly. Moreover,

communication fails when the LOS is blocked (i.e., the link is susceptible to shadowing). Additional drawback is that safety requirements are particularly stringent due to the deployment of directed sources.

A diffuse link operates entirely without LOS, which results in an increased robustness against shadowing and support of high user mobility within a large coverage area. Since no pointing or tracking is necessary, diffuse link presents the most desirable configuration from the users point of view. This is why it evoked much interest from the research community. The Spectrix system [18], rendering 4 Mbit/s over the surface of 1000 square feet ($\approx 93 \text{ m}^2$) is an example of a commercial product. A diffuse infrared link was even foreseen in the original IEEE 802.11 standard as one physical layer option [19]. Nevertheless, these links suffer from high optical path loss (i.e., they require larger optical powers) and are seriously limited by the inter-symbol-interference due to multipath dispersion. Beside the power budget, the achievable transmission rate in diffuse links depends also on room properties (size, reflection coefficients of surfaces, etc.,...).

Various power-efficient modulation schemes (variants of pulse position modulation, PPM [20–22], and of pulse interval modulation [23, 24]) were proposed and compared to the performance of the simplest non-return-to-zero on-off keying (NRZ-OOK) links. Nevertheless, since they achieve an improvement in power-efficiency on account of the increased bandwidth requirement, they are particularly sensitive to multipath dispersion. To combat the effects of multipath propagation and improve link performance, different techniques have been considered. For one, deployment of various angle-diversity receivers (non-imaging and imaging²) was proposed. In combination with a multi-beam transmitter (consisting of multiple narrow beams pointing in different directions), a 70 Mbit/s experimental link was reported, [25]. Maximum likelihood sequence detection, as well as different equalization approaches (linear and decision feedback equalizer, DFE) were investigated with different modulation formats, and bit rates up to 50 Mbit/s were demonstrated with use of OOK and DFE in a $10 \times 10 \text{ m}^2$ room [26]. Moreover, several spread-spectrum techniques (e.g., sequence inversion keying, SIK, complementary SIK, and M-ary bi-orthogonal keying) were specially designed for optical systems and investigated in [27], [28] (with rather low data rates of about 1 Mbit/s).

It is often desired in practical applications to combine mobility of the diffuse and high-speed capability of the LOS link. The so-called multi-spot diffusing approach was proposed to improve power-efficiency of diffuse links by sending multiple moderate-width or collimated beams to the ceiling in different directions,

²A non-imaging angle-diversity receiver consists of multiple narrow-field-of-view modules, oriented in different directions and each with its own optic, which together have a large FOV. The imaging receiver (also called “fly-eye”) consists of an imaging optical concentrator with a segmented photodetector array placed in its focal plane.

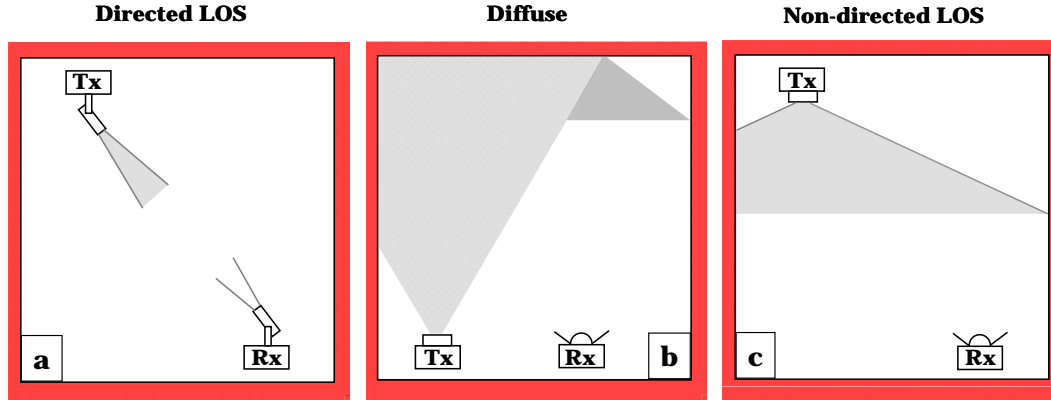


Figure 1.1: The basic designs of OW links.

where they would be diffusely reflected on their way to the angle-diversity receiver with non-imaging optics. Such link configuration relies on multiple lines-of-sight between the diffusing spots and the receiver. Different means to create multiple diffusing spots were proposed in [25] and [29]. In [30], the use of a self-orienting receiver has been proposed as an alternative to conventional angle diversity receivers. This receiver required an electromechanical orienting system, controlled by digital signal processing. The approach suggested in [31] considered to secure a directed link against connection failure by adding a diffuse mode (wide beam and large FOV) to an electronically tracked directed link (as in [17]). Where LOS connection is free, the directed link would allow high speeds (155 Mbit/s shown), and otherwise, the transceivers would switch to diffuse operation mode and the connection would be maintained over reflections at a fall-back rate. However, all these proposals consider rather complex multi-beam-forming transmitters and/or diversity-based receivers.

In order to benefit from the advantages of both connection types, alternatively a *non-directed LOS link* (shown in Fig. 1.1 c) is considered in this thesis. In such a link, LOS and diffuse signal components are *simultaneously* present at the receiver (assuming an uninterrupted LOS path). The equivalent channel response is characterized by high dynamics in both bandwidth and gain, depending above all, on LOS prominence. Measurements indicating such channel behavior are reported in e.g., [32], [33]. An often adopted approach to support the user mobility without outages is to realize the system design as statical, with transmission parameters fixed according to the worst case scenario. This is the simplest way to obtain a reliable connectivity under all channel conditions within the coverage area (for example, everywhere in the room). Even though simple, such a system cannot recognize or utilize the advantage of LOS presence (being that it is designed according to the purely diffuse channel condition), which results in an inefficient use of system (channel) capacity. In order to maximize the system throughput, while

reliable operation and full coverage are maintained, the system could be designed bandwidth- and rate-adaptive. An adaptive system reduces the data rate under the adverse channel conditions until a desired error performance is reached. The necessary channel state information is provided from the receiver over a reliable low-speed feed-back link. Since the adaptive behavior is realized with digital signal processing, it is possible to use simple and low-cost optical components.

The “time-domain” approach for rate-adaptive links considers serial transmission. Various proposals, based on adaptive coding with pulse position modulation and OOK were proposed in [34–38]. Combining such adaptive transmission with diversity techniques was considered in [39], and with a self-orienting receiver in [40]. However, baseband modulation schemes suggested in these works use wideband signals and, thus, suffer from inter-symbol-interference (ISI) due to multipath propagation. Therefore, complex time-domain equalization is needed at the receiver.

1.2 Goal and contributions of the thesis

Differently to the mentioned “time-domain” approach, this thesis considers a rate-adaptive system based on multiple-subcarrier transmission, where the aggregate system transmission rate is adjusted to the channel conditions by choosing different modulation and/or coding schemes for individual subcarriers. With such “frequency-domain” approach we can deal with both channel dynamics and multipath dispersion at the same time. Basic research on OW multiple subcarrier modulation (MSM) systems was conducted in [41–43]. It is known from the radio domain that multicarrier modulation (MCM)³ techniques (such as OFDM) deal inherently with multipath distortion, and permit flat-fading equalization (reduced complexity) at the receiver, [44]. Adaptive modulation technique is well known for electrical transmission in digital subscriber line (DSL) systems (referred to as discrete multitone, e.g., [45], [46]), and is currently under discussion for radio channels (under the name adaptive OFDM, AOFDM e.g., [47]). As mentioned above, specific properties of the OW channel pose a first moment constraint and non-negativity constraint on the transmit signal. These constraints demand an appropriate redesign of the adaptive OFDM technique as well as new performance evaluation. Unlike in the mobile radio channel, where the whole base-band is generally usable for transmission (except for sporadic fading holes in the spectrum), under poor conditions in the OW channel only a few subcarriers in the low-frequency range are useful. This means that even though the coded OFDM is very powerful in radio links (e.g., [48]), it is likely not to work for OW links and that the adaptive OFDM presents a better suited approach.

³Multicarrier modulation and multiple subcarrier modulation are in essence the same techniques, the only difference is in terminology. While the former is customary for radio systems, the latter is used in OW systems.

The overall goal of this thesis is to provide an insight on the potentials of the adaptive modulation technique in an OW OFDM-based system for a typical indoor WLAN scenario. It is one of the first, if not the first work to recognize the attractiveness of this technique. In that context, two main questions are handled

Where are the theoretical limits for transmission rate?

Finding the capacity of an IM/DD channel is a difficult fundamental problem which is unlikely to be solved in general. In this thesis, it is shown that the capacity analysis of RF OFDM systems can not be adopted directly for an OW OFDM system because of the different constraints in the channel. To that end, this thesis provides a detailed discussion on consequences of capacity analysis by a conventional method suitable for RF OFDM systems. As the main contribution, the upper capacity bound of a system with complex multiple subcarrier modulation and fixed bias in a frequency-selective optical IM/DD channel is derived.

How and to what extent can these theoretical limits be approached in practice?

Throughput maximization and rate adaptation in a practical system whose subcarriers are modulated by square M-ary quadrature amplitude modulation (M-QAM) and whose performance is additionally limited by an allowed error rate is achieved by deployment of so-called loading algorithms. This thesis shows that in order to strictly fulfil the conditions of the OW channel (i.e., signal non-negativity), a loading algorithm needs to comply with a rather conservative constraint (more stringent than the actual constraint imposed by the physical channel), and investigates how an algorithm which performs optimally in power-limited systems needs to be adjusted for optimal performance under this new constraint. The results show that a dynamically adaptive system provides great transmission rate enhancements compared to the statically designed one even under such a conservative condition. The upper capacity bound for such *conservative* adaptive M-QAM based system design (with strict non-negativity of the transmit signal waveform) is also derived.⁴

Furthermore, this thesis investigates the influence of controlled clipping (of the transmit signal waveform in front of the optical source) on the system performance in terms of both error and transmission rates. This approach considers a relaxed non-negativity constraint. The results show that the capacity bound of the system with complex multiple subcarrier modulation can be even more approached by tolerating a minor increase in error rate due to clipping. A theoretical capacity upper bound of the adaptive M-QAM based system design *which includes clipping*

⁴An M-QAM based system is a special case of the system with complex multiple subcarrier modulation. The upper capacity bound of the former is hence different (lower) than the capacity bound of latter system.

is also determined. It is shown that the adaptive modulation technique promises a possibility for reliable transmission rates of several hundreds of Mbit/s in the scenario considered in this thesis, even when using very simple optical front ends.

In the context of assessing the practicability of the investigated approach, supporting questions such as: *What is the adequate channel model and which set of model parameters is realistic and suitable for the communication scenario considered?*; *How should the simple opto-electrical front-ends of the system be designed in order to allow broadband transmission with sufficient power budget?*; *How often do we need to update the system transmission parameters for reliable connectivity?* are also thoroughly handled.

1.3 Outline of the thesis

This thesis consists of five technical chapters, with main contributions located in Chapters 4 to 6. In the following, the content of all chapters is briefly outlined. For the sake of clearer presentation, a more extended outline is given at the beginning of each technical chapter.

Chapter 2 introduces the communication scenario representing a bidirectional single input single output (SISO) link for WLAN application. A short review of a simple channel model suitable for non-directed LOS links is made. The influence of relevant parameters is discussed with the goal to obtain a set of values which provide a good agreement of the model and realistic conditions. Moreover, this chapter reviews the fundamentals of propagation and noise effects in the OW channel.

Chapter 3 describes the concept of a rate-adaptive system based on OFDM as a promising approach to handle channel dynamics and signal degradation due to multipath dispersion. Mathematical modeling of multipath channels is presented together with fundamentals of multicarrier transmission and OFDM as its special case. Functional blocks of an OW OFDM-based rate-adaptive system are introduced and differences to the conventional systems considered for radio applications are pointed out and discussed.

As a prerequisite for investigation of the transmission technique in focus, Chapter 4 analyzes and defines the opto-electrical front-ends of the system. The aim is to keep the component design simple and to allow broadband transmission with enough link budget at the same time. The system parameters obtained in this way are then used to support the numerical examples in the subsequent chapters.

Chapter 5 handles the evaluation of system capacity, according to the principles of information theory. In Section 5.1, capacity is defined, and the well-known expres-

sions for capacity of flat (narrowband) and frequency-selective (wideband) power- and band-limited Gaussian channels are presented. In this section, constraints on the input signal of the OW system are introduced. Section 5.2 evaluates the capacity of the system under these OW-specific constraints. In [49], You and Kahn derived an upper-bound for the capacity of optical IM/DD channels with complex multiple subcarrier modulation and fixed bias. The result is valid for a *flat* optical channel with high signal-to-noise ratio (SNR). By generalization of this result, in this thesis, an upper capacity bound of the system in a *frequency-selective* OW channel is derived. Given that the capacity analysis suitable for power-limited systems is well-known, in Section 5.3, we explain the compromises and relaxations of the initial capacity problem which are needed for proper application of this method in an OW system, and discuss the resultant estimation of system capacity.

In Chapter 6, the focus of investigations shifts to the throughput maximization in a practical system (where the subcarriers are modulated by square M-ary quadrature amplitude modulation (M-QAM) and performance is additionally constrained by a permitted error rate). The goal is to show the potential of efficient signal processing in the considered communication scenario. To that end, Section 6.1 considers throughput maximization via deployment of a loading algorithm which allocates modulation format and power to each subcarrier depending on the current channel state. Furthermore, permitting clipping of the signal in front of the optical source is investigated in Section 6.2 in terms of both error and transmission performance. Section 6.3 investigates the time interval which is needed for the update of system parameters (called system coherence interval, SCI) and discusses the consequences of the OW channel variation on system design.

Chapter 7 summarizes the main results and offers an outlook for future work.

Chapter 2

Communication Scenario and Channel

Section 2.1 introduces the communication scenario representing a bidirectional single input single output (SISO) link for WLAN application. A short review of a simple channel model, suitable for non-directed LOS links, is made in Section 2.3. A set of model parameters is chosen to provide a good agreement of the model and realistic conditions and, at the same time, to diminish the free space loss and provide a large coverage with a single access point unit. Investigations show that the OW channel strongly varies in the considered scenario both in magnitude and bandwidth, depending on the ratio of LOS and reflections' influences. Moreover, this chapter reviews on the fundamentals of propagation (Section 2.2) and noise effects (Section 2.4) in the OW channel.

2.1 Indoor communication scenario

As previously introduced, generic link types (purely LOS and purely diffuse link) offer complementary attractive qualities. In this thesis, it is considered to take the advantage of both the high-speed potential of the LOS link and high spatial coverage of the diffuse one, by regarding the *non-directed LOS* link, where both LOS and diffuse signals are simultaneously present at the receiver.

A typical example of a scenario where the non-directed LOS link could be of use is illustrated in Fig. 2.1. It represents communication between a fixed access point (Tx) and a mobile terminal unit, e.g., a laptop (Rx) in a moderate-size office space. Numerical simulations performed throughout this work assume the model room from Fig. 2.1 with an Rx which may be anywhere at the desk-top surface (distance r to the Tx), and which has a free orientation (parameter ψ at the Fig. 2.1 right), but the work is applicable for WLAN scenarios in general. The transmitter is assumed to be placed at the center of the ceiling and pointing downwards. Transmitter and receiver are assumed to have a wide-beam radiation characteristic and

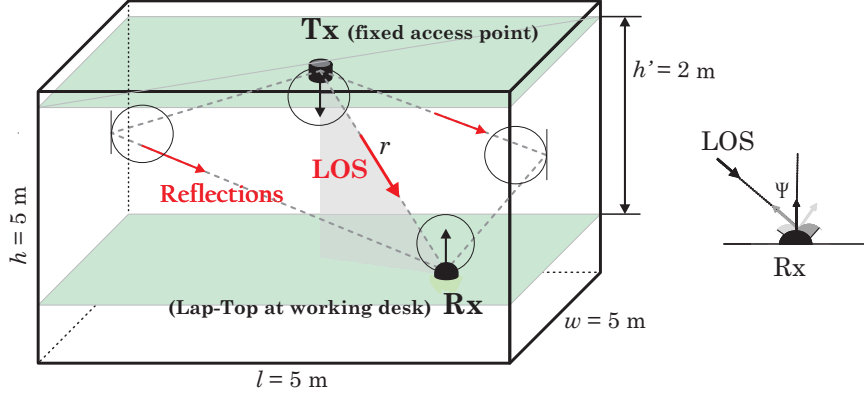


Figure 2.1: Communication scenario.

a wide field-of-view, respectively. Since no pointing or tracking mechanisms are foreseen, this makes a sensible choice for good coverage provision. It should also be noted that, even though in this work the access point at the ceiling is regarded as the transmitter and the mobile end-terminal as the receiver, the communication scenario is in general bidirectional and conclusions are valid for both directions.

2.2 Propagation effects in the OW channel

On their way from the transmitter to the receiver in a wireless non-directed link, optical waves traverse different paths, and therefore, experience different loss (gain) and propagation time (delay), as well as different angle of incidence to the receiver surface. Because of multipath propagation, the received signal is a superposition of differently delayed and attenuated replica of the transmitted signal, which undergo constructive and destructive interference when arriving at the Rx. The resultant effect of the received signal amplitude variation in time and space is in OW termed as *multipath fading*, whereas the signal spreading in time domain is referred to as *multipath dispersion*.

Differently than in radio, multipath fading is averaged out in OW communications. Multipath propagation causes the received electric field to suffer amplitude fades on the scale of wavelength. With direct detection technique, the photodetector generates a current proportional to the instantaneous optical power, that is, square of electric field. The total photocurrent is given by integration of the squared electric field (varying in amplitude and phase) over the entire detector surface. Thus, the short wavelength of the optical carrier and the huge relative size of the square-law detector (typically $10^3 - 10^6$ square wavelengths) lead to efficient spatial diversity, which makes the influence of multipath fading negligible.

Multipath dispersion is, however, very much present in OW, causing inter-symbol-interference when the pulse is smeared across more than one bit interval. Because it emerges from the superposition of different optical spatial modes, it can be considered as a linear channel effect. The influence of ISI in a multipath channel is quantified in time-domain by the channel *delay spread*, D . The delay spread is defined as the square root of the second central moment of the channel delay power spectrum $P_h(\xi) = E\{|h(\xi)|^2\}$ (also called the multipath intensity profile) [50]. For the optical wireless channel (where $P_h(\xi) = h^2(\xi)$), the delay spread is given as [51]

$$D = \left[\frac{\int_{-\infty}^{\infty} (\xi - \mu)^2 h^2(\xi) d\xi}{\int_{-\infty}^{\infty} h^2(\xi) d\xi} \right]^{1/2}, \quad \text{where } \mu = \frac{\int_{-\infty}^{\infty} \xi h^2(\xi) d\xi}{\int_{-\infty}^{\infty} h^2(\xi) d\xi} \quad (2.1)$$

is the mean delay (defined as the first central moment of the channel delay power spectrum). To avoid multipath dispersion, the delay spread should be significantly shorter than the symbol period $D \ll T_{\text{sym}}$. In [41] was stated that dispersion becomes a major degradation factor at transmission speeds about and above 10 Mbit/s.

Alternatively, the effects of multipath dispersion can be described in frequency domain over the so-called *channel coherence bandwidth* B_C [52], [53]. This bandwidth refers to the frequency range over which the channel affects adjacent frequency components in the same manner. The relationship between B_C and D is given as

$$B_C \approx \frac{1}{kD}, \quad (2.2)$$

where factor k takes values 1, 2, 2π and even 50, depending on the literature source. This relationship is intended as an order of magnitude relation, essentially pointing out that the coherence bandwidth is reciprocal to the multipath spread. Hence, in an equivalent way it can be stated that, in order to avoid multiple dispersion, the signal bandwidth needs to be significantly smaller than the channel coherence bandwidth $B \ll B_C$. In this case, all signal frequency components are considered to be affected in the same manner, which is referred to as *flat fading*. If, however, $B > B_C$, some frequency components of the signal are affected differently by the channel. Such *frequency selectivity* of the channel leads to the signal spread in time domain and ISI. Flat fading is the desired condition, since then no complex compensation (equalization) is required. In [51, 54, 55], delay spreads of about several tens of ns were measured or simulated for the non-directed channel in a medium-size room, rendering $B_C \approx 10$ MHz, which makes the OW channel in case of broadband transmission (with baseband signals of about 100 MHz) frequency-selective.

As long as the transmitter, the receiver and reflectors remain stationary, the OW channel impulse response is fixed, and D as well as B_C can be considered deterministic. In general, when these objects are moving relative to each other, the

frequency of the received signal will not be the same as at the source due to the Doppler effect. When the objects are approaching each other, the frequency of the received signal is higher, and when they are moving away from each other, it is lower than at the source. The *Doppler spread* B_D is defined as the Doppler shift at its maximum value

$$B_D = f_{\text{source}} \frac{v}{c}, \quad (2.3)$$

where f_{source} is the source frequency, v is the relative speed of the objects, and c the speed of light. The Doppler effect is described also by *channel coherence time* T_C , during which the channel effects on the signal may be considered as constant. Similarly as (2.2), in an order of magnitude sense, we can write

$$T_C \approx \frac{1}{kB_D}. \quad (2.4)$$

Again the relationship is somewhat imprecise, but the important thing is to recognize that the major effect in determining time coherence is the Doppler spread, and that the relationship is reciprocal; the larger the Doppler spread, the smaller the time coherence. If the channel characteristics change between the beginning and end of the symbol period, i.e., if $T_C < T_{\text{sym}}$, channel is considered as *rapidly time-varying*. As a result, the signal spreads in the frequency domain. A first order criterion to avoid the distortion due to frequency dispersion thus is $T_{\text{sym}} < T_C$, in which case the channel is considered as *slowly time-varying*. The larger T_C , relative to T_{sym} , the less significant the effect of Doppler spread to the receiver performance.

In principle, the channel significantly changes after the time of $T_C \approx c/(f_{\text{source}}v)$. Hereby, again due to direct detection, in OW, only the Doppler effect on the modulating signal is of relevance. Considering the typical speeds of objects moving indoors and typical aimed transmission rates (modulating frequencies), OW channel can be considered as invariant on short time scales.

Therefore, the equivalent baseband model of a multipath (non-directed) OW channel is described as a linear time-invariant (LTI) system, whose influence on the transmitted signal can be represented by linear convolution [56]. If $x(t)$ denotes the instantaneous electrical current in front of the optical source (modulating signal) in the noiseless IM/DD system, the received signal $y(t)$, i.e., the instantaneous photocurrent after detection, is given by

$$y(t) = \eta S x(t) * h(t) = \eta S \int_{-\infty}^{\infty} x(\tau) h(t - \tau) d\tau, \quad (2.5)$$

where $h(t)$ presents the impulse response of the equivalent baseband OW channel, and $S[\text{W/A}]$ and $\eta[\text{A/W}]$ the conversion factors at the optical source and the receiver, respectively. The operator $*$ denotes linear convolution. The influence of the OW channel h is therefore defined between the optical powers at the transmitter and receiver.

2.3 Noiseless OW channel model

A generally accepted analytical model for the OW channel of a non-directed LOS link, adopted in this thesis, was developed by Pohl et al. [55, 57]. This model provides a good rule-of-thumb prediction of essential channel properties, such as path loss and bandwidth. It is simple and yet takes into account an infinite number of reflections, unlike some other numerical models which considered, e.g., furniture or windows in the room, but suffered from high numerical complexity (limiting the accounted number of reflections to a few) [58], [59]. This section provides a short description of the analytical model and a discussion on relevant parameters, with the aim to obtain a set of favorite values to be used in investigations which follow.

In general, the impulse response of the non-directed LOS channel consists of two distinctive components - a discrete Dirac-like pulse (LOS contribution) and a continuous signal, arriving at the receiver after some time ΔT (contribution of diffuse reflections). It was recognized in [57] that the response of the diffuse channel is similar to that of an integrating sphere and that it can be modeled well with an exponentially decaying function, even though the real channel exhibits some peaks in the response (standing out of the exponential decay pattern) related to the first couple of reflections. A sketch of the channel model impulse response is depicted in Fig. 2.2. Analytically, it is a superposition of the LOS and diffuse channel responses

$$\begin{aligned} h(t) &= h_{\text{LOS}}(t) + h_{\text{DIFF}}(t - \Delta T) \\ &= \eta_{\text{LOS}}\delta(t) + \frac{\eta_{\text{DIFF}}}{\tau} e^{-\frac{t-\Delta T}{\tau}} u(t - \Delta T), \end{aligned} \quad (2.6)$$

where ΔT presents the delay between the LOS signal and first reflection at the receiver, η_{LOS} and η_{DIFF} represent path loss (gain) of the LOS and diffuse signals respectively, $1/\tau$ is the decay constant, and $u(t)$ is the Heaviside step function.

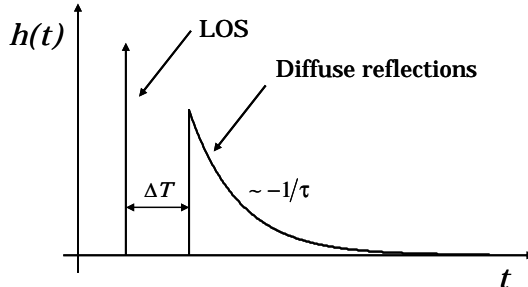


Figure 2.2: Sketch of the channel impulse response according to the adopted model.

Accordingly, the channel frequency response is a superposition of a flat and a

first-order low-pass transfer function

$$\begin{aligned} H(f) &= \eta_{\text{LOS}} + H_{\text{DIFF}}(f)e^{-j2\pi f\Delta T} \\ &= \eta_{\text{LOS}} + \eta_{\text{DIFF}} \frac{e^{-j2\pi f\Delta T}}{1 + j\frac{f}{f_0}}, \end{aligned} \quad (2.7)$$

where $f_0 = (2\pi\tau)^{-1}$ represents the 3 dB cut-off frequency of the diffuse channel. The delay of the diffuse component ΔT causes a frequency dependent phase offset $-2\pi f\Delta T$ in (2.7).

2.3.1 LOS component

The LOS response is practically independent of the modulating frequency (see (2.7)), and the LOS path gain depends on the geometry between transmitter and receiver. Path loss η_{LOS} is a function of transmitter radiation characteristic, distance between Tx and Rx (i.e., their positions in the room) and their orientations with respect to the LOS.

The radiation pattern of a wide-beam optical source, being either an LED or a laser with a diffusor is defined as the angular distribution of its output (optical) power P_{Tx} . It is modeled according to the generalized Lambert law [41], [60]

$$I(\phi) = \frac{dP_{\text{Tx}}}{d\Omega} = I_0 \cos^m \phi = P_{\text{Tx}} \frac{m+1}{2\pi} \cos^m \phi, \quad (2.8)$$

where $I(\phi)$ presents the radiant intensity in the direction ϕ , $I_0 = I(\phi = 0)$ presents the direction of maximum radiation, m the Lambert radiation index, which describes the shape of radiation characteristic, and $d\Omega$ is the solid angle covered by $I(\phi)$. The normalized shape of the generalized Lambert radiation pattern is illustrated in Fig. 2.3.

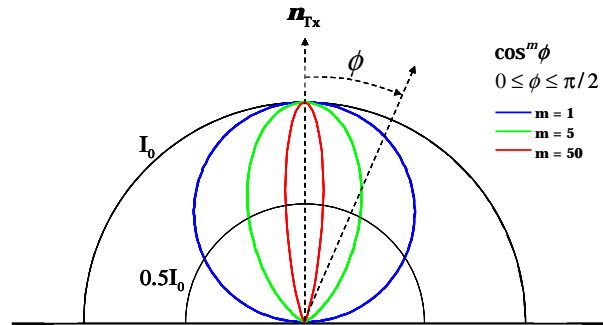


Figure 2.3: Normalized shape of the generalized Lambert radiation pattern.

Figure 2.4 illustrates source and receiver geometry with all relevant parameters. A receiver of surface A_{Rx} is placed at the distance r from the source. Directions of

maximum power radiation at the Tx (i.e., Tx orientation), maximum sensitivity of the Rx (i.e., Rx orientation) and of the LOS, are denoted by unit vectors \mathbf{n}_{Tx} , \mathbf{n}_{Rx} and \mathbf{n}_{LOS} , respectively. Then, the received power P_{Rx} can be obtained by multiplication of the radiant intensity (2.8) with the solid angle $\Omega = A_{\text{Rx}} \langle \mathbf{n}_{\text{Rx}}, -\mathbf{n}_{\text{LOS}} \rangle / r^2$ (assuming $A_{\text{Rx}} \langle \mathbf{n}_{\text{Rx}}, -\mathbf{n}_{\text{LOS}} \rangle \ll r^2$) covered by the receiver

$$P_{\text{Rx}} = P_{\text{Tx}} \frac{m+1}{2\pi} \langle \mathbf{n}_{\text{Tx}}, \mathbf{n}_{\text{LOS}} \rangle^m \frac{A_{\text{Rx}}}{r^2} \langle \mathbf{n}_{\text{Rx}}, -\mathbf{n}_{\text{LOS}} \rangle. \quad (2.9)$$

The representation $\langle \mathbf{n}_i, \mathbf{n}_j \rangle$ is used for scalar product. If both $\langle \mathbf{n}_{\text{Tx}}, \mathbf{n}_{\text{LOS}} \rangle$ and $\langle \mathbf{n}_{\text{Rx}}, -\mathbf{n}_{\text{LOS}} \rangle$ are positive, the LOS path exists and the LOS path gain can be simplified to

$$\eta_{\text{LOS}} = \frac{P_{\text{Rx}}}{P_{\text{Tx}}} = \frac{A_{\text{Rx}}}{r^2} \frac{m+1}{2\pi} \cos^m \phi \cos \psi. \quad (2.10)$$

Otherwise, $\eta_{\text{LOS}} = 0$. The receiver surface projected in the direction of the incoming radiation can be referred to as the detector effective surface $A_{\text{eff}}(\psi) = A_{\text{Rx}} \cos \psi$.¹

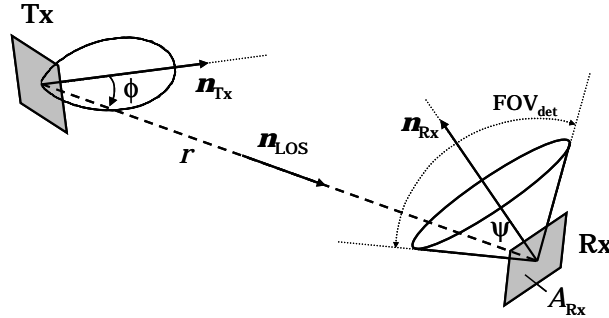


Figure 2.4: Geometry of source (Tx) and detector (Rx) without intermediate reflectors.

Lambert radiation index m is related to the so-called radiation semi-angle $\phi_{1/2}$ for which the maximum radiant intensity I_0 decreases to $I_0/2$

$$\phi_{1/2} = \arccos \sqrt[m]{0.5}. \quad (2.11)$$

Clearly, a higher Lambert index corresponds to higher directivity of the source radiation. For instance, for $m = 1$, the maximum intensity reduces by half in a direction $\phi_{1/2} = 60^\circ$ away from the axes, whereas for $m = 2$ and $m = 3$ the same occurs already by 45° and 37.5° , respectively. From Fig. 2.3, $\phi_{1/2}$ for $m = 5$ and $m = 50$ can be estimated. On the other hand, a higher m also implies higher

¹In Chapter 4.2, it shall be seen that the actual effective surface of the optical wireless receiver includes influences of an optical concentrator and optical filter. However, in this chapter, we speak only of a bare detector as receiver.

maximum intensity $I_0 = P_{\text{Tx}}(m+1)/2\pi$ (see (2.8)). A Lambert source, therefore, exhibits a trade-off between intensity and directivity.

Figure 2.5 shows the behavior of a normalized η_{LOS} for several parameters m , when the distance between the transmitter and receiver planes remains constant ($h' = \text{const.}$, see Fig. 2.1). After introducing $r = h'/\cos\phi$ in (2.10), η_{LOS} is normalized, making the plots independent of receiver surface and orientation, and of h' . Furthermore, $I_0 = 1$ in case $m = 1$ is assumed. From Fig. 2.5, it can be observed that LOS power is higher around the radiation axes for sources with higher directivity, but it falls more rapidly as the radiation angle grows. Therefore, there is a trade-off between the LOS power and coverage (size of the LOS cell). The choice of Lambert index m should therefore result in a strong LOS, but also in a good coverage of the whole Rx-plane in the room model from Fig. 2.1, or of the whole hot-spot area, if such is desired in the room. The inset in Fig. 2.5 shows that the maximum radiation angle ϕ_{max} , possible in the considered room scenario (with dimensions as in Fig. 2.1), when the LOS signal can reach the Rx, is $\phi_{\text{max}} = \arctan(\sqrt{l^2 + w^2}/(2h')) \approx \pi/3$. It is the case when Rx is placed in the corner of the room at the desk-top surface. Obviously, for the coverage of the whole desk-top surface using one source (as is foreseen in this thesis), the Lambert index can not be too high. The cases of $m = 1, 2$ and 3 are, therefore, chosen for further consideration.

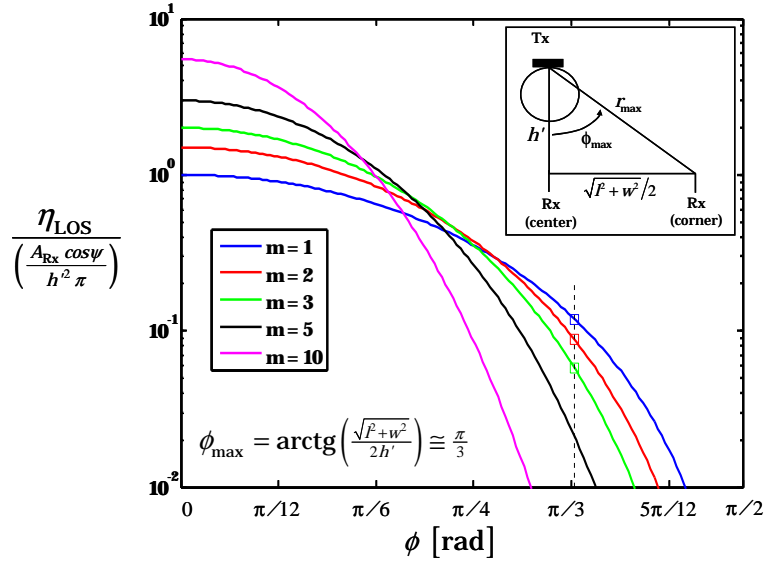


Figure 2.5: Influence of the Lambert radiation index on the LOS signal gain. The inset shows that in the room model as in Fig. 2.1, the maximum radiation angle of interest ϕ_{max} (corresponding to the receiver placed in the corner) is about $\pi/3$.

2.3.2 Diffuse component

As stated by (2.7), the contribution of the reflections in the room is approximated by the first-order low-pass transfer function (alike response of an integrating sphere). A detailed derivation of the impulse response for the sphere is given in [61]. The diffuse path gain (loss) η_{DIFF} was derived in [55], assuming that the first diffuse reflection of a wide-beam optical source creates homogenous intensity over the entire room surface A_{Room} , and that contributions of the following infinite number of reflections each further reduce this intensity by $\langle\rho\rangle$, where $\langle\rho\rangle$ presents the average reflectivity of all surfaces in the room. The received power of the diffuse signal was obtained by multiplying the effective receiver surface with the homogenous intensity after an infinite number of reflections, which finally rendered the diffuse signal path gain of

$$\eta_{\text{DIFF}} = \frac{P_{\text{Rx}}}{P_{\text{Tx}}} = \frac{A_{\text{Rx}} \sin^2(\text{FOV}_{\text{det}}/2)}{A_{\text{Room}}} \frac{\langle\rho\rangle}{1 - \langle\rho\rangle}. \quad (2.12)$$

The influence of the detector field-of-view (FOV_{det} in Fig. 2.4) on the received power is in (2.12) accounted for by the factor $\sin^2(\text{FOV}_{\text{det}}/2)$.

The reciprocal value of the exponential decay constant τ , determining the diffuse channel cut-off frequency f_0 , is given by

$$\tau = -\frac{\langle t \rangle}{\ln\langle\rho\rangle}, \quad (2.13)$$

where $\langle t \rangle$ is the average time between two reflections and the term $-1/\ln\langle\rho\rangle$ denotes the average number of reflections which are likely to occur until a photon is absorbed. In [55], it was shown that obtaining $\langle t \rangle$ like in a sphere, but with parameters of a rectangular space

$$\langle t \rangle = \frac{4V_{\text{Room}}}{cA_{\text{Room}}} = \frac{2lwh}{c(lw + lh + wh)}, \quad (2.14)$$

is an excellent approximation for rooms with similar dimensions l , w , h (denoting the room length, width and height, respectively as in Fig. 2.1), where V_{Room} stands for the room volume and c is the speed of light. Measurements of the channel frequency response, performed by Jungnickel et al. in an empty laboratory (with windows and floors covered with wallpaper so that reflectivity is more homogeneous), demonstrated an excellent agreement with the sphere model up to the cut-off frequency [55].

According to the assumed model, the size of the confined space influences the diffuse channel characteristic $H_{\text{DIFF}}(f)$ (2.7) both in terms of gain and cut-off frequency. Figure 2.6 shows the amplitude of diffuse channel characteristic, for three

different sets of room dimensions. Beside the theoretical predictions of the adopted model ($H_{\text{DIFF}}(f)$ in (2.7)), in each case, results of ray tracing simulations are given, showing a good agreement with the model. All simulation parameters can be found in Table A.1 in Appendix A. One can observe that, in a larger room, the diffuse signal gain decreases since the probability, that the signal reaches the receiver (whose size remains constant) after a number of reflections, decreases. Moreover, in larger rooms the average time between two reflections increases, leading to the lower cut-off frequencies, which is also observable in Fig. 2.6. The discrepancy from the model comes at high frequencies, for which the first reflections are responsible. At low frequencies, the agreement for the shown cases is very good. Slight discrepancy in the η_{DIFF} is due to the rectangular room instead of a sphere. This is why for the rooms where one dimension differs a lot than the other two (corridor), the sphere model is not the best choice. This was investigated in more detail in [57].

It has to be noted here that because of ray tracing simulations an unrealistically large detector surface is assumed (diameter of 10 cm) in order to decrease the simulation time. The resultant diffuse channel gain shown at the y -axis in Fig. 2.6 is hence also unrealistically large. This is however of minor relevance since the purpose of the figure is to show the agreement of simulation results with the analytical model. The same is valid for all other figures where results of ray tracing are presented. Realistic gains can be obtained simply by scaling $\eta_{\text{DIFF,actual}} = (A_{\text{Rx}}/78.5 \text{ cm}^2)\eta_{\text{DIFF}}$, when a detector of realistic surface A_{Rx} is assumed.

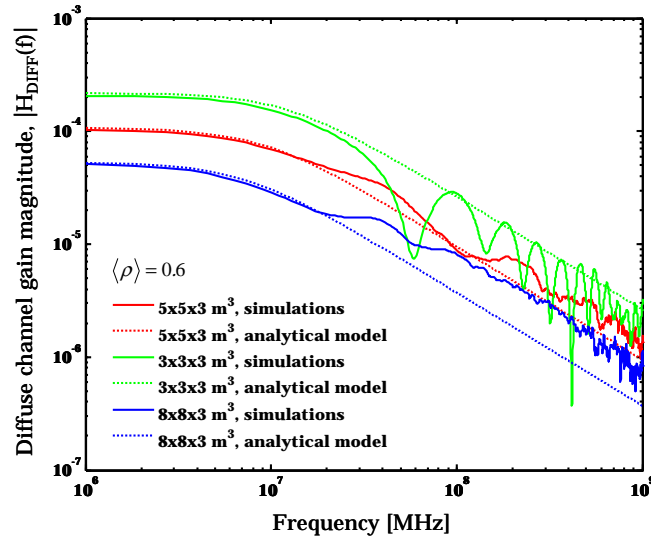


Figure 2.6: Influence of room dimensions on the diffuse channel characteristic. Simulation and calculation parameters are given in Table A.1 in Appendix A. The analytical model is given by H_{DIFF} from (2.7) with η_{DIFF} obtained by (2.12). Cut-off frequency $f_0 = (2\pi\tau)^{-1}$ is obtained by combining (2.13) and (2.14) for each room.

The influence of average reflectivity on the diffuse channel characteristic is also two-fold. Namely, with larger $\langle\rho\rangle$, the portion of radiated light which reaches the receiver (i.e., which is not absorbed before) is larger, and the resultant diffuse channel gain is higher. At the same time, the average number of reflections $-1/\ln\langle\rho\rangle$ increases, leading to a decline of the cut-off frequency. These effects are shown in Fig. 2.7, for three different average reflectivity parameters. Again, theory predictions are supported by simulations. The values of relevant parameters are given in the column corresponding to the $5 \times 5 \times 3 \text{ m}^3$ room, in Table A.1 of Appendix A.

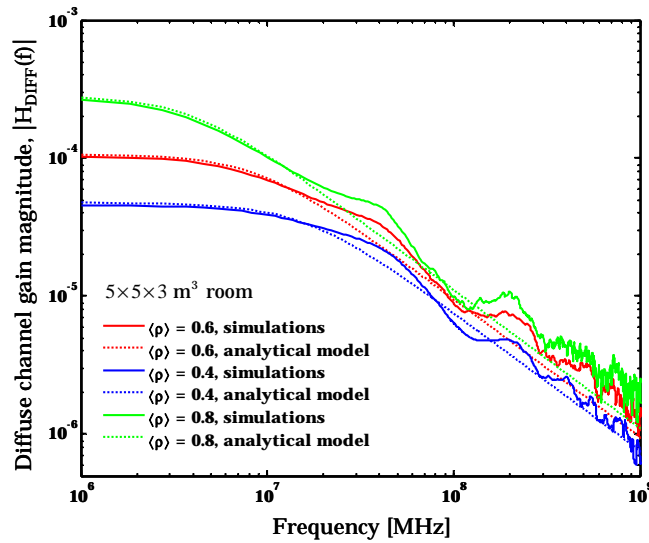


Figure 2.7: Influence of average reflectivity of the surfaces in the room on the diffuse channel characteristic. Simulation and calculation parameters are listed in Table A.1 in Appendix A.

According to the channel model, variation of receiver size and field-of-view both scale the gain η_{DIFF} , without influencing the cut-off frequency. Effect of FOV_{det} variation is verified by simulations and shown in Fig. 2.8. Gain curves have been normalized to the maximum gain (at zero frequency) with the $\text{FOV}_{\text{det}}/2 = 90^\circ$. Simulation parameters can be found in the column corresponding to the $5 \times 5 \times 3 \text{ m}^3$ room, in Table A.1 of Appendix A. It is clear that the best channel gain is obtained with a flat receiver, since it allows apprehension of the largest portion of light. For the same reason, also the receiver size should be maximized. Nevertheless, it cannot be chosen freely since it is limited by the desired photodetector bandwidth. This question will be addressed in Chapter 4.

In all further investigations, a $5 \times 5 \times 3 \text{ m}^3$ - room with an average reflectivity of $\langle\rho\rangle = 0.6$ and a flat receiver are assumed. Such parameters are chosen because they could very well reflect realistic circumstances. For example, if it is assumed

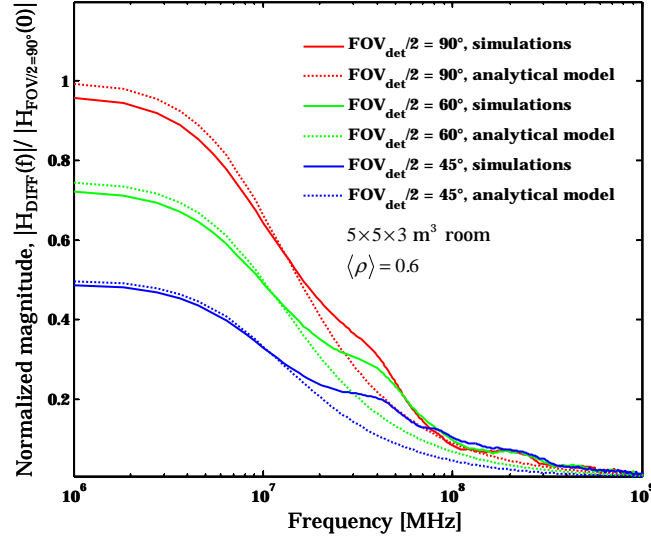


Figure 2.8: Influence of receiver FOV on the diffuse channel characteristic. Simulation and calculation parameters are listed in Table A.1 in Appendix A.

that an empty $5 \times 5 \times 3 \text{ m}^3$ - room has a window instead of one wall and a dark floor (both with $\rho \approx 0$), while the other surfaces (walls) are white ($\rho \approx 1$), the resultant average reflectivity is $\langle \rho \rangle \approx 0.63$. Additionally, from the previously mentioned measurements, in the empty laboratory, an average reflectivity of over 0.7 was estimated [55]. Moreover, the adopted channel model is well justified for this set of parameters.

2.3.3 Composite (LOS plus diffuse) channel

The frequency response of the adopted non-directed OW channel model is given by (2.7). Diffuse channel cut-off frequency f_0 is, as previously discussed, determined by the room properties. The two parameters which determine the total channel response are the phase offset $2\pi f \Delta T$ and the so-called K -factor defined in [55] as ratio of the (*electrical*) LOS and diffuse channel gain. The association of the K -factor with electrical power ratio at the output of the photodetector was intentional, in order to accommodate easier comparison with the Rician K -factor, which is well known in radio communications. In that context,

$$K[\text{dB}] = 20 \log \frac{\eta_{\text{LOS}}}{\eta_{\text{DIFF}}}. \quad (2.15)$$

In the cases when LOS is blocked, $K[\text{dB}] = -\infty$.

The phase offset is determined by the delay of the diffuse signal component with respect to the LOS, ΔT . Clearly, in the room model from Fig. 2.1, where Tx is fixed at the center of the ceiling, and Rx is at the desk-top surface, the delay ΔT

will be maximal for the Rx directly under the Tx, and minimal when Rx is placed against the wall (or in the corner). By ray tracing simulations, impulse responses are obtained for several Rx positions at the desk-top diagonal, including the two extreme cases (center and corner). The delay ΔT is calculated as $\Delta T = \mu_{\text{DIFF}} - T_{\text{LOS}}$, where μ_{DIFF} denotes the mean delay of the diffuse channel response, as given by (2.1). Figure 2.9 shows the values for ΔT in these cases, ranging from 0 to some 20 ns.

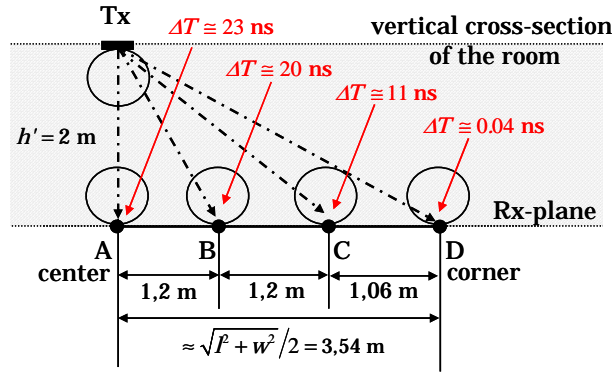


Figure 2.9: Variation of diffuse component delay depending on Rx position on the desk-top plane.

Figure 2.10 shows the composite channel frequency response magnitude $|H(f)|$ obtained from the analytical model (2.7) for several illustrative K -factors. Channel gain is calculated over the 300 MHz frequency range and assuming $\Delta T = 10$ ns. It is clear that the channel response highly depends on LOS prominence (described by the K -factor). Where the LOS is blocked or is very weak, the channel response is approximately low-pass and its bandwidth is quite small. As LOS gets more pronounced, the channel response varies until it becomes almost flat for sufficiently large K -factors, rendering bandwidths by an order of magnitude greater than in the diffuse case. The notches in the channel characteristic appear due to destructive interference of the two components at frequencies f when the phase condition

$$\arg(H_{\text{LOS}}) - \arg(H_{\text{DIFF}}(f)) = 2\pi f \Delta T + \arctan \frac{f}{f_0} = n\pi, \quad n = 1, 3, \dots \quad (2.16)$$

is satisfied, but are especially deep when also the amplitude condition

$$|H_{\text{LOS}}| = |H_{\text{DIFF}}(f)| \quad (2.17)$$

is fulfilled. According to (2.16), the notches are shifted towards smaller frequencies for larger ΔT . This is shown in the Fig. 2.11, where the intercepts of the plots indicate the frequency of the first notch ($n = 1$ in (2.16)) for $f_0 \approx 9$ MHz and three cases of ΔT . In the figure, blue plots are the result of $\pi - 2\pi f \Delta T$ and the red plot

of $\arctan f/f_0$. For example, in the Fig. 2.10 for $K = -10$ dB, the first notch (on frequency $f \approx 30$ MHz) is especially deep, since the crossing point between the LOS and diffuse component characteristic is in the vicinity. For smaller K -factors, the amplitude condition for the deep notch shifts towards the second, third, etc. notch frequency.

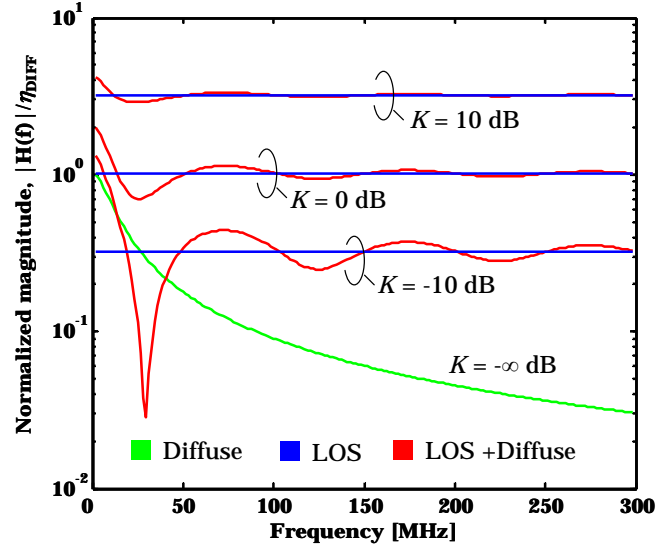


Figure 2.10: Magnitude of the optical wireless channel frequency response (2.7) for different values of the K -factor. Diffuse channel gain is obtained by (2.12) and $\Delta T = 10$ ns is assumed.

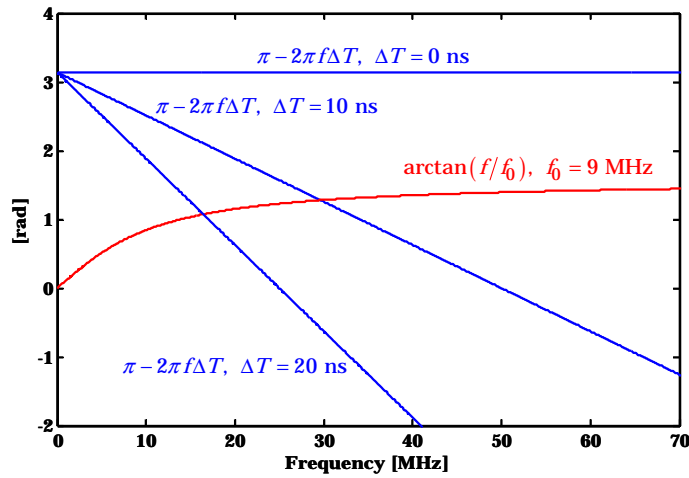


Figure 2.11: Dependence of the first notch frequency on ΔT . The intercepts of curves (reflecting (2.16)) render the frequency of the first notch if $n = 1$.

In further investigations, it is intended to use the K -factor as a single parameter to describe the channel state. In order to obtain the span of K -factors of interest, distribution of the K -factor was calculated across the Rx-surface (desk-top height) in the room from Fig. 2.1, for three values of Lambert index m . Figure 2.12 shows the distribution of K -factor values, depending on the parameter $r \sin \phi$ (horizontal Tx-Rx distance, see Fig. 2.1) and the angle ψ (angle between the LOS arrival direction and Rx orientation direction in Fig. 2.4).² It is assumed that LOS is not blocked.

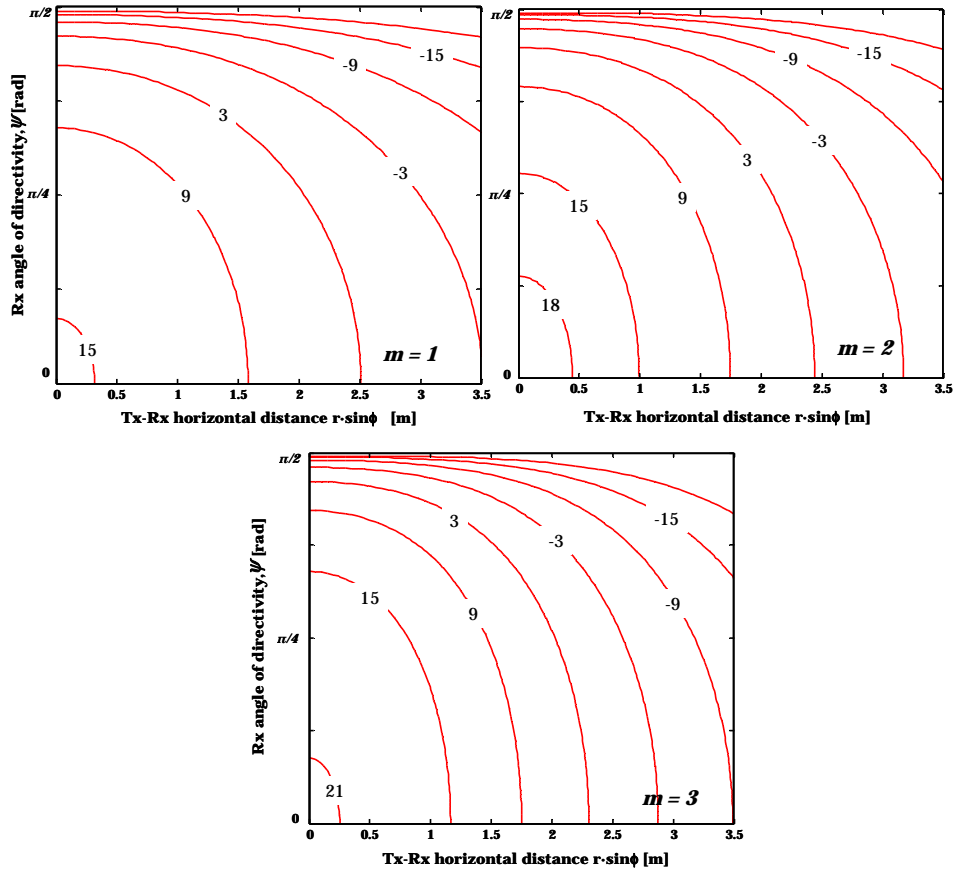


Figure 2.12: Distribution of the K -factor in [dB] (contour plots) at the desk-top plane, depending on the Rx orientation ψ and its horizontal distance from Tx, $r \sin \phi$.

Clearly, a source of higher directivity results in larger values of K near the center of the room. This is the reason why even for worse Rx orientation, better K -factor values can be obtained with a more directed source. At the same time, for sources

²These parameters are sufficient to describe the distribution of the K -factor, because of the symmetry in the assumed model room. Nevertheless, since only a span of K -factors is of interest, there is no loss of generality.

with higher m , K -factors decrease more rapidly with the increase of distance (i.e., the LOS loses power faster) and the “hot-spot” area is smaller. Nevertheless, since the case $m = 3$ results in K -factors better than 0 dB in the area up to 2.5 m from the Rx-plane center, which is realistic for a “working area” hot-spot, in the following investigations a K -factor span $[-20, +25]$ dB is considered.

As already indicated in Fig. 2.10, the channel bandwidth varies with LOS prominence. If the 3-dB cut-off frequency is defined by

$$|H(f_{3\text{dB}})|^2 = \frac{|H(f=0)|^2}{2}, \quad (2.18)$$

one can calculate the electrical channel bandwidth offered in cases of different K -factors present at the Rx-surface. The results are shown in Fig. 2.13, where $\Delta T = 10$ ns is assumed. It can be observed that with LOS stronger than the diffuse component ($K > 0$ dB), the channel bandwidth starts to grow significantly, first quite steadily, then however, followed by a sudden increase of more than an order of magnitude (with respect to the purely diffuse channel), for K -factors greater than 7 dB. For some poor channels ($K < 3$ dB), the available bandwidth actually turns out to be slightly smaller than in the purely diffuse case. This happens when the K is such that the amplitude condition for the deep notch is fulfilled at the very small frequencies (so the first notch in the characteristic is the deep one). This notch causes the amplitude characteristic to have a steep decrease. The channel characteristic shown in Fig. 2.10, when $K = -10$ dB is a good example. Nevertheless, the channel magnitude, which is at the lowest frequencies better than for the purely diffuse channel, ensures better transmission conditions (i.e., higher capacity) than in the diffuse case even for such K -factors. (This is observable in Chapters 5 and 6.)

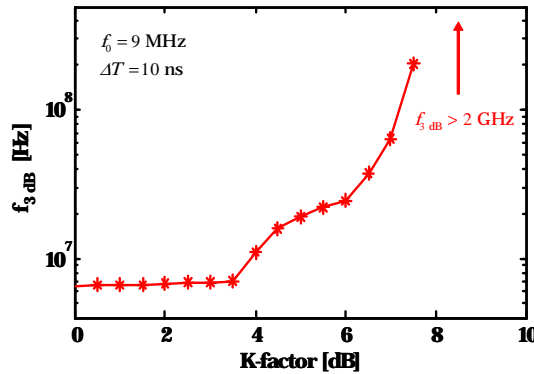


Figure 2.13: Dependence of channel 3dB frequency on K -factor.

2.4 Noise sources and mechanisms in the OW channel

The most important source of noise in the OW channel is the background light. This is not the case in the fiber-optic channel (where the noise is a result of the receiver amplifier circuit), nor is it the case in the conventional radio wireless channel (where most of the noise is coming from other users).

Generally, the background light presents a mixture of radiation coming from the sun, as the natural source, and incandescent and fluorescent lighting, as artificial sources. These sources radiate power over a broad range of wavelengths. To their maxima normalized, the spectral power densities of different sources are shown in Fig. 2.14 together with the normalized sensitivity curve of a silicon photodiode (Si-PD) [41], [60]. Also shown are the radiation band of a near-IR laser diode (LD) and of a light-emitting diode (LED).

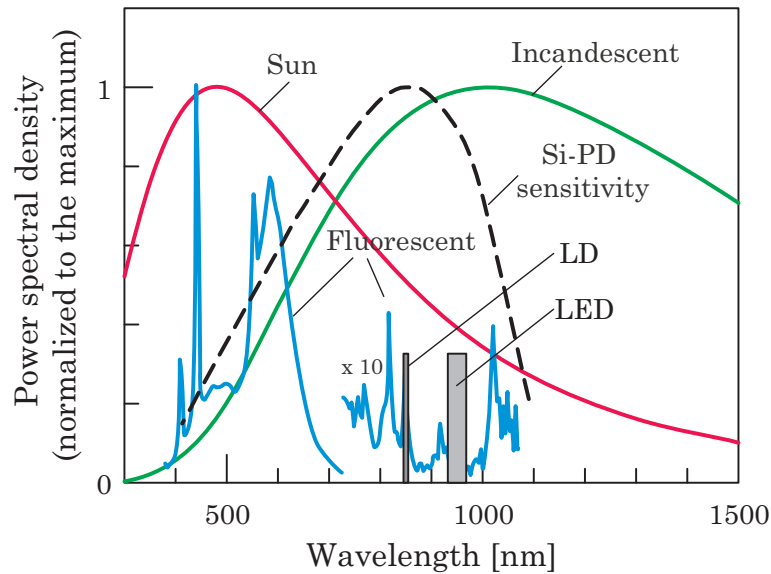


Figure 2.14: Coincidence of background radiation with Si-photodiode responsivity [41], [60]. All plots are normalized to their maxima.

The natural light is practically stationary (variations of the light intensity due to, e.g., shadowing are very slow), whereas the light produced by artificial sources exhibits large and fast fluctuations in time. Therefore, both natural and artificial sources produce *shot noise* on the photodiode, but fluorescent artificial sources additionally generate an interference at frequencies of hundreds of kilohertz [62], making such lamps potentially a much more serious impairment (depending on the modulation scheme).

Shot noise is a consequence of the generation of electrons of the photocurrent at random times.³ When a signal of constant optical power P_{Rx} arrives at the photodetector with responsivity η , the generated photocurrent is

$$I(t) = \eta P_{\text{Rx}} + i(t), \quad (2.19)$$

where $\eta P_{\text{Rx}} = I_{\text{pc}}$ is the average photocurrent, and $i(t)$ is the current fluctuation due to shot noise. At the receiver, a DC blocker can eliminate I_{pc} , however the power of the AC part (variance) remains and contributes to the noise in the link. This fluctuation is a stationary random process with Poisson statistics, which is in practice (due to high intensity) approximated with Gaussian statistics [64]. Shot noise is an example of white noise, with a spectral density proportional to the average current I_{pc} . The noise variance at the receiver with the effective noise bandwidth B (depending on the receiver design) is given by

$$\sigma^2 = 2qI_{\text{pc}}B, \quad (2.20)$$

where $q = 1.6 \cdot 10^{-19}$ [As] denotes the electron charge.

In literature, the most thorough investigations of the shot noise induced by the background light were presented in [60] and [62]. Based on the results (presented in terms of average background light photocurrent assuming a particular detector), it was concluded that direct sunlight produces the highest level of noise, which can be even some orders of magnitude greater than the noise in an artificially well-lit environment. This is the reason why a bright-sky-light spectral irradiance ($5.8 \mu\text{W}/\text{cm}^2\text{nm}$ [41], [60]⁴) is assumed as a basis for further investigations of receiver design conducted in Chapter 4. The spectral irradiance can be assumed as independent of wavelength within the bandwidth of the optical filter [51].

Emission of the fluorescent lamps lies at spectral lines of mercury and argon (780 – 950 nm) and is approximately periodically modulated at the frequencies of 50 or 60 Hz in case of lamps with “conventional ballast”, or at the frequencies of tens to hundreds of kHz in case of lamps with “electronic ballast”. Whereas the electrical spectrum of the former contains significant harmonics up to tens of kHz, the spectrum of the latter was measured to contain energy up to hundreds of kHz [62]. Such interference can present serious impairments for optical wireless links, especially in the case of base-band transmission. To ensure a certain

³Historically, the fluctuations of the current, flowing from the heated cathode to the anode of a vacuum tube, were named shot noise, due to the sound made by the audio amplifiers whose gain was such that these fluctuations were the dominant noise source. Later on, quantum theory of light has shown that photon arrival time at the photocathode of a photodetector exhibits the same Poissonian statistics as the thermionic emission of electrons from the hot cathode of a vacuum tube [63].

⁴In [41], this value presents the irradiance at 810 nm. For the sake of simplicity, it is also assumed in this thesis, since the actual value for the sources operating at higher wavelengths can only be slightly smaller, as observable in Fig. 2.14.

link performance, the needed signal power would depend on the applied base-band modulation scheme. On the other hand, subcarrier modulation could be made immune to fluorescent-light noise, by choosing the lowest subcarrier frequency higher than these harmonics.

In order to suppress the background noise, it is desired to use a narrow-band optical source (such as laser) and a narrow pass-band optical filter before detection. The actual average photocurrent coming from background radiation will depend, therefore, on the receiver design, especially of its optical front-end components (optical filter and concentrator).

Precisely speaking, the total shot noise at the receiver is actually the sum of several components resulting from influences of the transmitted signal, the dark current, the gate leakage current and the ambient light. However, in the case of typical infrared receivers, the first three components can be neglected [41]. Apart from the shot noise, *thermal noise* (generated at the load resistor of the receiver preamplifier) also contributes to the link performance degradation. The power of thermal noise depends on the preamplifier circuit design. A more detailed noise analysis with numerical examples is given in Chapter 4.2, where a suitable receiver design is discussed.

Chapter 3

OFDM-Based Rate-Adaptive OW System

Chapter 3 describes the concept of a rate-adaptive system based on OFDM as a promising approach to handle channel dynamics and signal degradation due to multipath dispersion. In Section 3.1, mathematical modeling of multipath channels is presented and fundamentals of multicarrier transmission and OFDM as its special case are reviewed in Sections 3.2 and 3.3. Functional blocks of an OW OFDM-based rate-adaptive system are introduced in Section 3.4 and differences to the conventional systems considered for radio applications are discussed.

At the beginning, it is convenient to define the notational conventions used in this chapter (but also later on). Distinction is made between signals in frequency domain, which are denoted by capital letters (like X), and signals in time domain, denoted by small letters (like x). Indices m and n are used for spectral signal components (f -domain), whereas indices k and i are used for the samples of discrete signals in time domain. Vectors or matrices are denoted in squared brackets by bold letters as $[\mathbf{x}]$ or $[\mathbf{X}]$, depending on the domain. Elements of vectors or matrices are non-bold-face letters (like $X(n)$ or X_n , which are both used depending on convenience).

3.1 Mathematical modeling of noiseless multipath channels

Although the physical channel ($h(t)$ in (2.5)) and (2.6)) is a continuous-time medium, due to digital signal processing considered in this thesis, it is of interest to use its discrete-time equivalent (sampled at the same instants of time as the signal). In that context, let the input signal $x(t)$ (from (2.5)) be bandwidth-limited to frequency B and sampled with the interval $T_{\text{sam}} = 1/2B$. The sampled output signal $y(k) = y(t = kT_{\text{sam}})$ results from $x(k) = x(t = kT_{\text{sam}})$ passing through a

noiseless¹ *discrete LTI channel* with impulse response h_D as [53]

$$y(k) = \sum_{i=-\infty}^{\infty} h_D(i)x(k-i). \quad (3.1)$$

The coefficients of the equivalent discrete LTI channel impulse response representation are given by $h_D(i) = T_{\text{sam}}h(t = iT_{\text{sam}})$. Because the discrete representation is considered throughout the rest of the thesis, for the sake of simplicity, the notation without the index D is adopted also for the discrete channel response. It is, however, not to be confused with values of the continuous channel response at the sampling instances. In (3.1) and until stated otherwise, the opto-electrical conversion parameters $S = 1$ [W/A] and $\eta = 1$ [A/W] are considered, without loss of generality.

Let the sampled channel have a finite impulse response no longer than L_C samples.² Its linear convolution with a discrete signal of length $N \geq L_C$ can be represented by a matrix-vector product

$$[\mathbf{y}] = [\mathbf{h}][\mathbf{x}] \quad (3.2)$$

with

$$[\mathbf{y}] = [y(0) \ y(1) \ \cdots \ y(N + L_C - 1)]^T, \quad (3.3)$$

$$[\mathbf{h}] = \begin{bmatrix} h(0) & 0 & \cdots & & & 0 \\ h(1) & h(0) & 0 & \cdots & & 0 \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ h(L_C - 1) & \ddots & \ddots & h(0) & 0 & \\ 0 & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & h(L_C - 1) & \ddots & \ddots & h(0) \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & & 0 & h(L_C - 1) & h(L_C - 2) \\ & & & & 0 & h(L_C - 1) \end{bmatrix},$$

$$[\mathbf{x}] = [x(0) \ x(1) \ \cdots \ x(N - 1)]^T.$$

The result is a vector of $N + L_C - 1$ elements. In (3.3), operation $[\cdot]^T$ denotes vector or matrix *transpose*.³ An LTI channel clearly stretches the input signal, hereby causing the distortion of the signal which is sent immediately afterwards

¹Since the purpose of this section is to characterize effects of multipath dispersion, the additive noise contribution is not considered before Section 3.4.

²The channel length is determined by dividing the maximum path delay (that is, the delay spread D defined in Section 2.2) with the sampling period T_{sam} .

³The transpose of a complex matrix is often defined to include conjugation of elements. The conjugating transpose operation is also called the Hermitian transpose. To avoid confusion, in this thesis, $[\cdot]^T$ and the word “transpose” always denote transposition without conjugation.

(or after a sufficiently short period of time). This type of distortion is in general referred to as inter-symbol interference (ISI). One possibility to handle ISI is to use adaptive filters for equalization at the receiver. This can be rather complicated because the number of filter taps required grows with the number of channel taps (i.e., with transmission rate).

3.2 Multicarrier transmission and OFDM

As already introduced in Chapter 2.2, a time-dispersive channel has the property of frequency selectivity in frequency domain. Figure 3.1 shows how a frequency-selective channel affects the signal of bandwidth B in baseband (blue shaded illustrations). At the channel output, different frequencies occupied by the signal are differently influenced by the channel and signal reconstruction requires a rather complex equalization in the time domain.

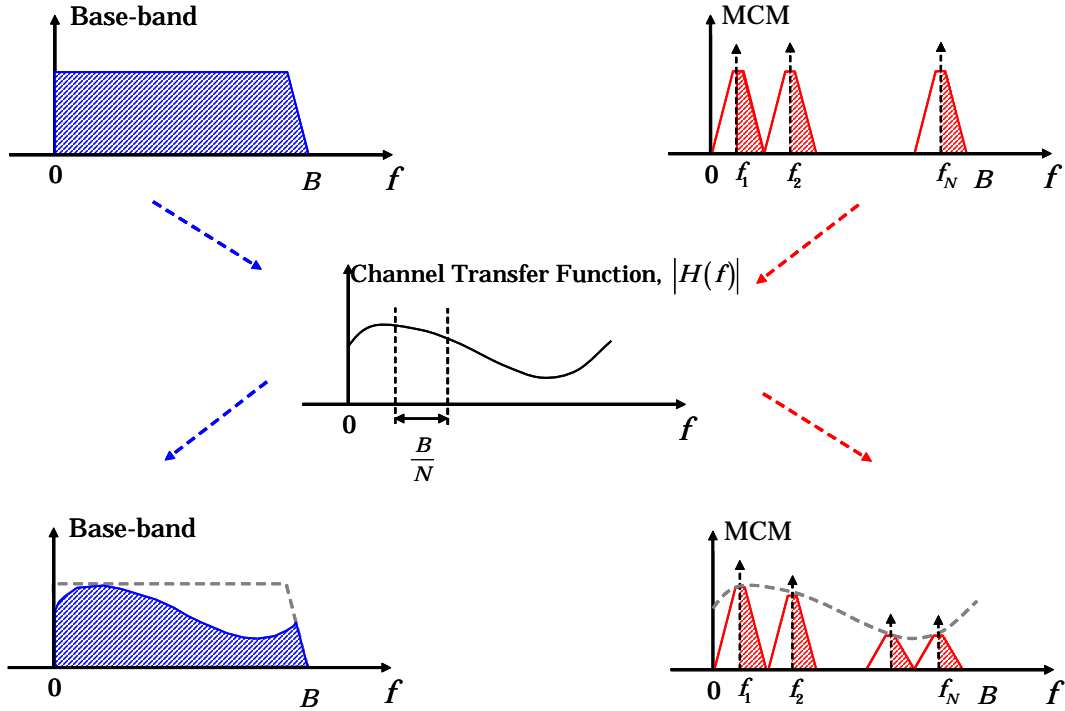


Figure 3.1: Passing of baseband (shaded blue) and MCM (shaded red) signals through a frequency selective channel. Channel characteristic is merely an illustration.

Equalization at the receiver can be made much simpler and more efficient if the bandwidth B occupied by the signal is divided into smaller sub-bands. Then, a part of the information is allotted to each sub-band and transmitted in parallel. This approach is referred to as *multicarrier modulation* (MCM) and it is presented in Fig. 3.1 by red shaded illustrations. If it can be considered that, when passing

through these relatively narrow frequency bands, signals experience flat-fading in the subchannels, then the equalization (reconstruction) of the signal in each sub-band requires a filter with only one tap.

Figure 3.2 shows how multicarrier modulation is realized, i.e., how the transmission of a broadband signal over a frequency-selective channel can be turned into a transmission of several narrowband signals over flat-fading channels. The input data stream is split into several (N) parallel streams, each of which modulates a different frequency (termed carrier or subcarrier) f_1, \dots, f_N .

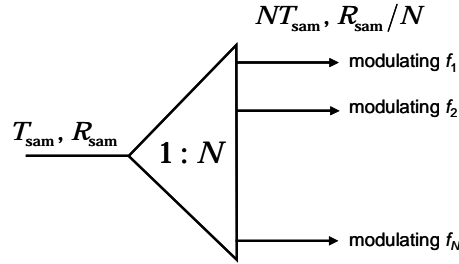
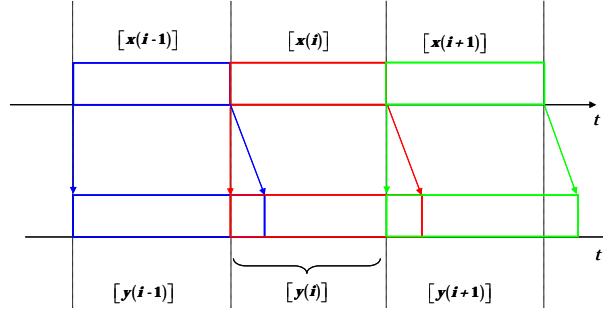


Figure 3.2: Parallel data transmission, T_{sam} and R_{sam} denote signaling interval and rate of the original (serial) data stream.

Let the original data stream occupy a bandwidth B , i.e., let it have a sampling (signaling) interval of $T_{\text{sam}} = 1/2B$. After an $1:N$ serial-to-parallel conversion, the parallel streams have an N times lower signaling rate (R_{sam}/N) than the original data stream, the signals on the subcarriers occupy relatively narrow frequency bands in baseband (B/N). For a sufficiently large N , flat fading can be assumed for the parallel signal streams. Equivalently, the duration of the symbol interval on parallel streams is N times longer (NT_{sam}). The delay spread hence corrupts a smaller percentage of the subcarrier symbol period. By multicarrier modulation, N samples of the input signal stream are transmitted in parallel. It makes therefore sense to speak of *block-wise* transmission, where one block consists of N samples and has a duration of NT_{sam} .

Transmitting the data block-wise and in parallel on several (many) carriers is very useful for mitigating the effects of time-dispersive media. With the channel L_C samples long, let now the signal samples be grouped into blocks of length $P \gg L_C$ which are transmitted immediately one after another. From (3.3) it is clear that the channel stretches input signal. Due to the channel influence, each block affects the first $L_C - 1$ samples of the next block, and itself suffers the influence of the previous block in its first $L_C - 1$ samples. One can hence speak of *inter-block interference* that arises due to dispersive nature of the channel. The i^{th} block at the channel output $[\mathbf{y}(i)]$ depends on both $[\mathbf{x}(i)]$ and $[\mathbf{x}(i - 1)]$ blocks at the channel input. This is illustrated in Fig. 3.3.

**Figure 3.3:** Illustration of inter-block-interference.

Mathematically, we can separate the influences of blocks $[\mathbf{x}(i-1)]$ and $[\mathbf{x}(i)]$ by writing [65]

$$[\mathbf{y}(i)] = [\mathbf{h}_0] [\mathbf{x}(i)] + [\mathbf{h}_1] [\mathbf{x}(i-1)], \quad (3.4)$$

where $[\mathbf{x}(i)]$, $[\mathbf{x}(i-1)]$ and $[\mathbf{y}(i)]$ are vectors of length P , and $[\mathbf{h}_0]$ and $[\mathbf{h}_1]$ are $P \times P$ matrices, defined as

$$[\mathbf{h}_0] = \begin{bmatrix} h(0) & 0 & \cdots & & 0 \\ h(1) & h(0) & 0 & & \vdots \\ \vdots & & \ddots & \ddots & \\ h(L_C - 1) & & & h(0) & 0 \\ 0 & h(L_C - 1) & & h(0) & \ddots \\ \vdots & & \ddots & & \ddots & 0 \\ 0 & \cdots & 0 & h(L_C - 1) & \cdots & h(1) & h(0) \end{bmatrix}, \quad (3.5)$$

$$[\mathbf{h}_1] = \begin{bmatrix} 0 & \cdots & 0 & h(L_C - 1) & \cdots & \cdots & h(1) \\ \vdots & & 0 & h(L_C - 1) & & & h(2) \\ & & & \ddots & \ddots & \vdots & \\ & & & 0 & h(L_C - 1) & & \\ & & & & 0 & \vdots & \\ & & & & & 0 & \vdots & \\ 0 & \cdots & & & & & 0 \end{bmatrix}.$$

Matrices $[\mathbf{h}_0]$ and $[\mathbf{h}_1]$ are parts of the matrix $[\mathbf{h}]$ from (3.3).⁴ From (3.4) and Fig. 3.3 it is clear that the influence between consecutive blocks can be eliminated if a guard interval (at least L_C long) is inserted between information samples. In that way blocks can be made independent of each other and can be processed separately. Throughout the rest of the section, it is hence assumed that a transmission

⁴Matrix $[\mathbf{h}_1]$ has additional $P - (L_C - 1)$ “zero” rows.

block of P samples consists of N information and $L \geq L_C$ guard samples. The redundancy introduced by the guard band leads to a decreased spectral efficiency $N/(N + L)$, however with a relatively large N (number of subcarriers in an MCM system), good spectral efficiencies can be achieved.

From $[\mathbf{h}_0]$ (i.e., from the first summand in (3.4)) it is obvious that a dispersive channel also causes ISI between samples within one block. The data within one block are transmitted in parallel over different frequencies so, in order to distinguish it from inter-block-interference, this type of interference is in an MCM system usually referred to as *inter-carrier interference*. In the classical MCM approach, which is often referred to as *frequency division multiplex* (FDM), inter-carrier interference is avoided by choosing subcarriers with disjunctive spectra, so that undesired components can be filtered out. Unless a complex single side-band modulation is considered, the FDM approach has in the best case the same spectral efficiency as baseband modulation.

Spectral efficiency of FDM can be improved if the subcarriers are chosen as mutually orthogonal. Two complex functions are said to be orthogonal if, when one is multiplied with the conjugated other and integrated over the domain of interest, the integral becomes zero [53]. In general, subcarriers are represented as complex exponentials, and the waveforms of two subcarriers on frequencies f_n and f_k are orthogonal on the time domain \mathcal{D} if

$$\int_{t \in \mathcal{D}} e^{j2\pi f_n t} e^{-j2\pi f_m t} dt = \begin{cases} 1, & \text{for } n = m, \\ 0, & \text{otherwise.} \end{cases} \quad (3.6)$$

For a symbol duration T_{sym} , orthogonal subcarriers occupy multiples of the lowest frequency $f_1 = 1/T_{\text{sym}}$. For instance, if f_m and f_n are multiples of $1/T_{\text{sym}}$, according to (3.6) they are orthogonal on the domain $\mathcal{D} = [0, T_{\text{sym}}]$. If orthogonal, subcarriers are allowed to have overlapping bandwidths without resulting in inter-carrier interference. In the technique generally known as *orthogonal frequency division multiplex* (OFDM), adjacent subcarriers have a minimum spacing possible ($\Delta f = 1/T_{\text{sym}}$).

Figure 3.4 shows how bandwidth efficiency of OFDM with respect to FDM grows with the number of subcarriers N . It is assumed that the information in baseband occupies bandwidth B .⁵ Whereas the needed transmission bandwidth with FDM always remains $2B$ (unless a signal side-band modulation is considered), in an

⁵Actually, in a practical system, pulse shaping is done with a raised-cosine filter to limit the influence outside the pulse interval. As a consequence, the effective bandwidth increases from B to $B(1 + \alpha)$, where $\alpha \in [0, 1]$ is the so called roll-off factor of the raised-cosine pulse. With larger α pulse decay in time domain is faster, but the needed transmission bandwidth is also larger. In Fig. 3.4, factor $(1 + \alpha)$ was left out, since it applies to both the FDM and OFDM case, and does not influence the difference in spectral efficiency.

OFDM system it depends on N as $(N + 1)B/N$, i.e., it decreases with the number of subcarriers. For large N , the needed bandwidth approaches B . The advantage in efficiency is in the practical system somewhat decreased by insertion of the mentioned guard interval.

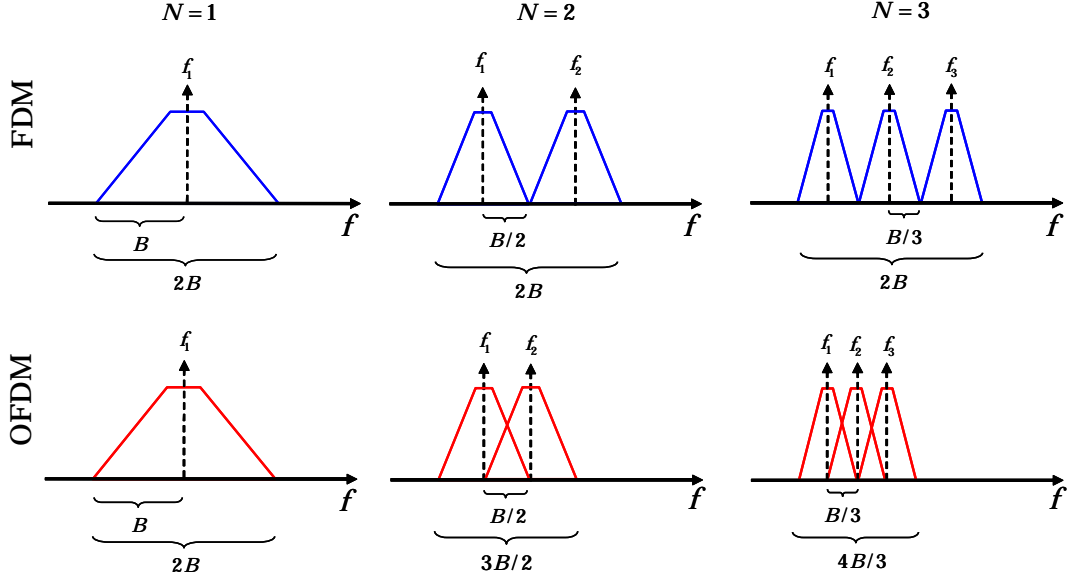


Figure 3.4: Comparison of spectral efficiency of FDM and OFDM. Carrier separation in OFDM is B/N .

In order to be able to distinguish among subcarriers at the receiver, preservation of their orthogonality throughout the transmission is of outmost importance in an OFDM system. This is achieved by designing the guard interval in a special way.

3.3 OFDM transmission chain

In 1960s, an OFDM modulator was first conceived and implemented with analog oscillators [47], [66], as presented in Fig. 3.5 a. However, to achieve resilience against fades in mobile wireless channels, a large number of orthogonal subcarrier modems (with base angular frequency $\omega_0 = 2\pi/T_{\text{sym}}$) was required. Because of this, OFDM did not achieve great popularity until its fully digital implementation was enabled by means of discrete Fourier transform, DFT, which has orthogonal basis functions. Implementation of fast Fourier transform (FFT) algorithm brought further significant hardware simplifications.

The DFT approach is illustrated in Fig. 3.5 b. By performing an N -point inverse DFT (IDFT) on the input data vector of complex-valued symbols $X(n) = a_n + jb_n$, $n = 0, \dots, N - 1$, the information is imprinted on N mutually orthogonal subcarriers $f_n = n/T_{\text{sym}}$. By definition, the IDFT transforms the input vector

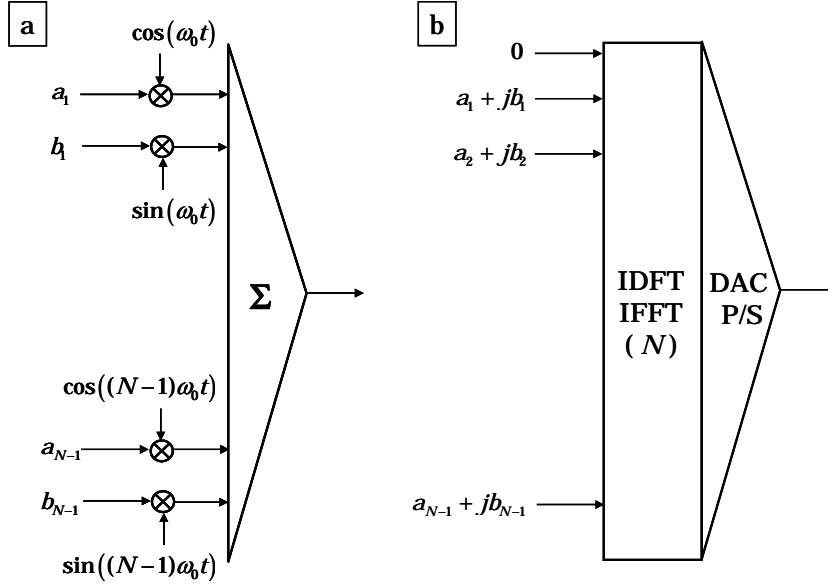


Figure 3.5: Multicarrier modulator in a) analog implementation and b) digital implementation. P/S denotes parallel-to-serial and DAC digital-to-analogue conversion.

$[\mathbf{X}]$ (with elements $X(n)$) from frequency-domain into a discrete output $[\mathbf{x}]$ in time-domain with samples

$$x(k) = \frac{1}{N} \sum_{n=0}^{N-1} X(n) e^{j2\pi nk/N}, \quad k = 0, \dots, N-1. \quad (3.7)$$

Since the first input in IFFT block (i.e., $n = 0$) corresponds to zero frequency, this “subcarrier” is usually left unmodulated. Because of this, only $N - 1$ modulated subcarriers are illustrated in Fig. 3.5 a. Similarly, OFDM demodulation (not shown in Fig. 3.5) is performed at the receiver by a DFT on the received OFDM block $[\mathbf{y}]$ in time-domain

$$Y(n) = \sum_{k=0}^{N-1} y(k) e^{-j2\pi nk/N}, \quad n = 0, \dots, N-1, \quad (3.8)$$

returning the signal back into the f -domain ($[\mathbf{Y}]$).

A simplified block scheme in Fig. 3.6 a illustrates relevant functional blocks and signal vectors in the OFDM transmission chain. Here, all the mentioned signal vectors $[\mathbf{x}]$, $[\mathbf{X}]$, $[\mathbf{y}]$ and $[\mathbf{Y}]$ are observable. To overcome the effects of the multi-path channel, a special guard interval is inserted in each OFDM symbol (block) in time domain (after IFFT). This guard band is formed simply by taking a number of samples (L) from the end of each symbol, and copying them as its prefix, as

shown in Fig. 3.6 b. It is hence referred to as *cyclic prefix* (CP).

In Appendix B, mathematical fundamentals leading to the system input-output relationship (between vectors $[\mathbf{X}]$ and $[\mathbf{Y}]$ in Fig. 3.6 a) are given. The derivation presented in Appendix B shows how insertion of a sufficiently long cyclic prefix after the IFFT at the transmitter (block ICP) and its removal before the FFT at the receiver (block RCP) converts a serial linear channel (based on linear convolution) to a system based on a circular convolution which is free of both inter-block- and inter-carrier-interference.

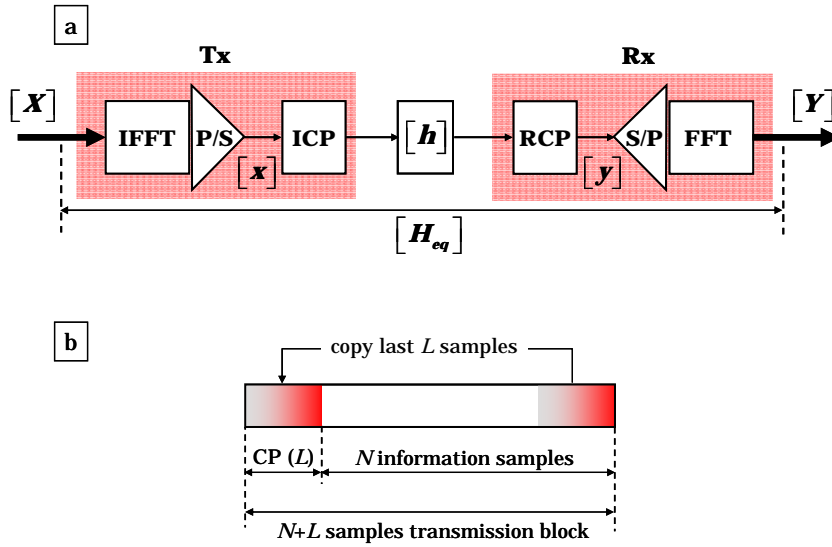


Figure 3.6: (a) Simplified scheme of block-wise transmission with DFT over an LTI channel, where ICP and RCP stand for insertion and removal of cyclic prefix respectively; (b) Generation of cyclic prefix guard interval.

In other words, the obtained system input-output relationship is linear

$$[\mathbf{Y}] = [\mathbf{H}_{eq}] [\mathbf{X}], \quad (3.9)$$

where $[\mathbf{X}]$ and $[\mathbf{Y}]$ denote sent and received vectors (in f -domain) of length N respectively, while \mathbf{H}_{eq} presents an $N \times N$ *diagonal* equivalent channel matrix. The elements on the main diagonal (i.e., the non-zero elements) of \mathbf{H}_{eq} present the channel frequency response on the subcarrier frequencies

$$H_{eq}(n, n) = H(n) = H(e^{j2\pi f_n}) = \sum_{k=0}^{L_C-1} h(k) e^{-j2\pi \frac{nk}{N}} \quad (3.10)$$

for $n = 0, \dots, N-1$. Taps of the discrete channel impulse response $h(k)$ in (3.10) are the same ones which were introduced by (3.1) and (3.3).

Multicarrier interpretation of the input/output relationship

As previously mentioned, a detailed step-wise derivation of the system input-output relationship (3.9) is presented in Appendix B. In the following passage, an interesting interpretation of (3.9) is given, which reveals the multicarrier nature of the OFDM system. The action of N -IFFT module on the input vector $[\mathbf{X}]$ with elements $X(n) = a_n + jb_n$ can be represented as

$$[\mathbf{x}] = \sum_{n=0}^{N-1} \frac{1}{N} [\mathbf{F}_n] X(n), \quad (3.11)$$

where $[\mathbf{x}]$ denotes the IFFT output vector whose elements are defined by (3.7) and $[\mathbf{F}_n]$ presents a column vector

$$[\mathbf{F}_n] = [e^{j0} \quad e^{j2\pi n/N} \quad e^{j2\pi 2n/N} \quad \dots \quad e^{j2\pi(N-1)n/N}]^T. \quad (3.12)$$

Vector $[\mathbf{F}_n]$ can be considered as a sampled version of the continuous carrier signal $e^{j2\pi f_n t}$, where $t \in [0, NT_{\text{sam}}]$, NT_{sam} being the length of an OFDM symbol without guard interval. Since $[\mathbf{F}_n]$ is a complex sinusoid of finite duration in time domain, its DFT is a *sinc* function. The spectrum of the vector $[\mathbf{x}]$ (i.e., the OFDM signal in frequency domain) is, therefore, a sum of N *sinc* functions, separated by $\Delta f = 1/(NT_{\text{sam}})$. This is shown for Fig. 3.7, where $N = 8$. It is hence intuitive to speak about subcarriers, $n = (0), 1, \dots, N - 1$ being the subcarrier index. As previously mentioned, frequency f_0 is usually left unmodulated, so only $N - 1 = 7$ *sinc* functions can be observed.

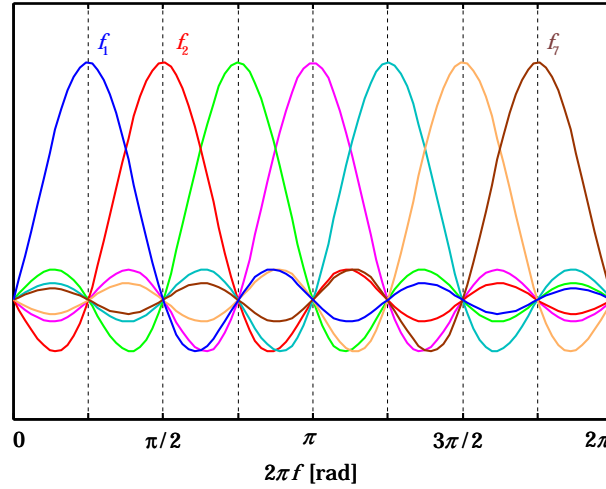


Figure 3.7: Spectrum of $N - 1 = 7$ equally modulated subcarriers in OFDM.

Complex exponentials are eigen-vectors of all LTI systems. This means that they pass through the system unchanged in shape, but only scaled with a certain eigen-

value, which presents the system frequency response. The orthogonality of eigenvectors after passing through the LTI system is also retained. In this manner, it can be shown that the channel (including the generation and removal of cyclic prefix, to obtain a linear system, see Appendix B) scales vectors $[\mathbf{F}_n]$ by the channel frequency response at the frequencies f_n . The signal vector at the channel output is therefore

$$[\mathbf{y}] = \sum_{n=0}^{N-1} H(e^{j2\pi n/N}) \frac{1}{N} [\mathbf{F}_n] X(n). \quad (3.13)$$

This illustrates how multicarrier techniques turn a convolutive (frequency-selective) channel into a superposition of N multiplicative (flat) channels. If $k = 0, \dots, N-1$ denotes the index of the time-domain sample, then the k^{th} element of $[\mathbf{y}]$ is

$$y(k) = \sum_{n=0}^{N-1} H(e^{j2\pi n/N}) \frac{1}{N} F_n(k) X(n), \quad (3.14)$$

where $F_n(k)$ presents the k^{th} element of the vector $[\mathbf{F}_n]$. After the FFT at the receiver, at the system output there is a vector

$$\begin{aligned} [\mathbf{Y}] &= \sum_{k=0}^{N-1} [\mathbf{F}_k]^{-1} y(k) \\ &= \sum_{k=0}^{N-1} [\mathbf{F}_k]^{-1} \left(\sum_{m=0}^{N-1} H(e^{j2\pi m/N}) \frac{1}{N} F_m(k) X(m) \right), \end{aligned} \quad (3.15)$$

with n^{th} subcarrier (element)

$$\begin{aligned} Y(n) &= \sum_{k=0}^{N-1} F_k^{-1}(n) \left(\sum_{m=0}^{N-1} H(e^{j2\pi m/N}) \frac{1}{N} F_m(k) X(m) \right) \\ &= \frac{1}{N} \sum_{m=0}^{N-1} H(e^{j2\pi m/N}) X(m) \left(\sum_{k=0}^{N-1} e^{j2\pi km/N} e^{-j2\pi kn/N} \right). \end{aligned} \quad (3.16)$$

The inner sum is equal to zero for all products except for $m = n$, which leaves

$$\begin{aligned} Y(n) &= \frac{1}{N} H(e^{j2\pi n/N}) X(n) N \\ &= H(e^{j2\pi n/N}) X(n), \end{aligned} \quad (3.17)$$

revealing again the linear nature of the input output relationship. For the whole system, we arrive again at the relationship (3.9).

Because of the linear input-output relationship, each element of the output vector depends only on the corresponding element of the input vector as $Y(n) =$

$H(n)X(n)$. This means that the simplest equalization is possible in an OFDM system. Such equalization is referred to as linear frequency-domain correction by division (or zero forcing, or channel inversion) since the symbol sent on each subcarrier can be reconstructed simply by division of the received symbol with the channel coefficient on the corresponding frequency

$$\left[\widehat{\mathbf{X}}\right] = [\mathbf{H}_{\text{eq}}]^{-1} [\mathbf{Y}]. \quad (3.18)$$

In the noiseless channel $\left[\widehat{\mathbf{X}}\right] = [\mathbf{X}]$.

3.4 Rate-adaptive OW system

Great advantage of OFDM modulation is its inherent ability to eliminate inter-symbol-interference. Nevertheless, a frequency selective channel influences signals on individual subcarriers with different loss factors, which can cause significant SNR decrease on some of them. In particular, the output on the n^{th} subcarrier (f -domain) in the i^{th} OFDM symbol (t -domain), with noise in the system, is

$$Y_n(i) = H_{\text{eq},n}X_n(i) + Z_n(i), \quad (3.19)$$

where $X_n(i)$ is the transmitted symbol, $H_{\text{eq},n}$ the coefficient of the equivalent channel matrix associated to the n^{th} subcarrier and $Z_n(i)$ is the spectral component of additive white gaussian noise (due to ambient light) on the n^{th} subcarrier. In this section, an OW system is in focus, where the equivalent channel includes also the influences of the O/E conversion factors at the optical source (S [W/A]) and receiver (η [A/W]), so we can write $H_{\text{eq},n} = \eta H_n S$. The electrical SNR on the subcarrier n is defined as

$$\text{SNR}_n = \frac{(\eta S)^2 \langle X_n^2 \rangle |H_n|^2}{\langle Z_n^2 \rangle} = \frac{P_n (\eta S)^2 |H_n|^2}{N_n}, \quad (3.20)$$

where the averaging is performed over the time of many OFDM symbols. Here, the electrical power of the transmitted signal is denoted with P_n and the noise power after detection in one subchannel with N_n . Although equalization at the receiver

$$(\eta S H_n)^{-1} Y_n(i) = X_n(i) + (\eta S H_n)^{-1} Z_n(i) \quad (3.21)$$

removes the impact of channel loss on the desired signal, it enhances the noise if $\eta S H_n < 1$. As any linear equalization, it does not change the SNRs, meaning that equalization alone will not improve transmission on poor subcarriers.

In Chapter 2, it is shown that the time-dispersive OW channel is characterized by great dynamics in bandwidth, depending above all on the LOS prominence. Depending on the Tx, room and Rx characteristics, the channel frequency response

varies from low-pass (about 10 MHz bandwidth) to approximately flat over the frequency range of hundreds of MHz.⁶ It is, therefore, difficult to maintain an SNR sufficient for high bit rates under all channel conditions, while using reasonable transmitter power levels. If, on one hand, the system is designed so that sufficient SNR is achieved on all subcarriers according to the worst-case (purely diffuse) channel characteristic, the baseband bandwidth offered under favorable channel conditions cannot be efficiently used, resulting in rather conservative transmission rates. On the other hand, if the system is designed with parameters fixed for transmission rates requiring bandwidths larger than offered by the diffuse channel, error performance becomes prohibitively poor, likely leading to system outage.

In order to exploit the channel bandwidth in an efficient way (i.e., to maximize the transmission rate, while maintaining reliable operation and large coverage), it is more appropriate to consider dynamically adjusting the set of used subcarriers, as well as their modulation formats, to the current channel condition.⁷ Thus, a good subcarrier (faced with a favorable channel gain and relatively small noise enhancement) would use a higher-order modulation scheme and carry more information, whereas a poorer one would carry less or even no information. At the same time, transmission with a required bit-error-rate (BER) can still be realized despite the impairments in the channel. In this way, such a *rate-adaptive* system is able to transmit at high speeds under favorable channel conditions, and to reduce throughput (until desired BER is achieved) as channel degrades. In order to follow the variations in the OW channel, adaptive transmission requires accurate channel estimates at the receiver and a reliable feed-back link between Rx and Tx, to provide the channel state information to the transmitter.

The functional block-scheme of an adaptive OFDM-based optical wireless system is given in the Fig. 3.8. At the Tx, the input data run in several parallel streams, which are appropriately mapped to a chosen modulation format or order (in the adaptive mapping block), and imprinted onto orthogonal subcarriers (in the IFFT block). The mapping of a stream is performed according to the information on the low-rate feedback link. In Fig. 3.8, the feed-back link carries a noise enhancement vector (coefficients for each subcarrier), based on which the loading algorithm determines power and modulation order of data on individual subcarriers. An OFDM symbol is then formed by digital signal processing. Samples in time domain are acquired by IFFT, after which a cyclic prefix of an appropriate length is added,

⁶In contrast to the case of broadband mobile radio channels, where more or less the whole baseband (except for sporadic fading holes in the spectrum) is usable for transmission, the available bandwidth in OW, i.e., number of usable subcarriers varies significantly (so that under the poor channel conditions only some sub-carriers in the low-frequency range are exploitable).

⁷Additionally, other parameters, like coding rate of error correction coding schemes can also be adapted to the channel condition (not considered in this thesis but could be a topic for further research).

as described in Section 3.3. After passing the samples through a digital-analog-converter (DAC) and low-pass filter (LPF), the continuous time-domain signal is used to directly modulate the optical source.

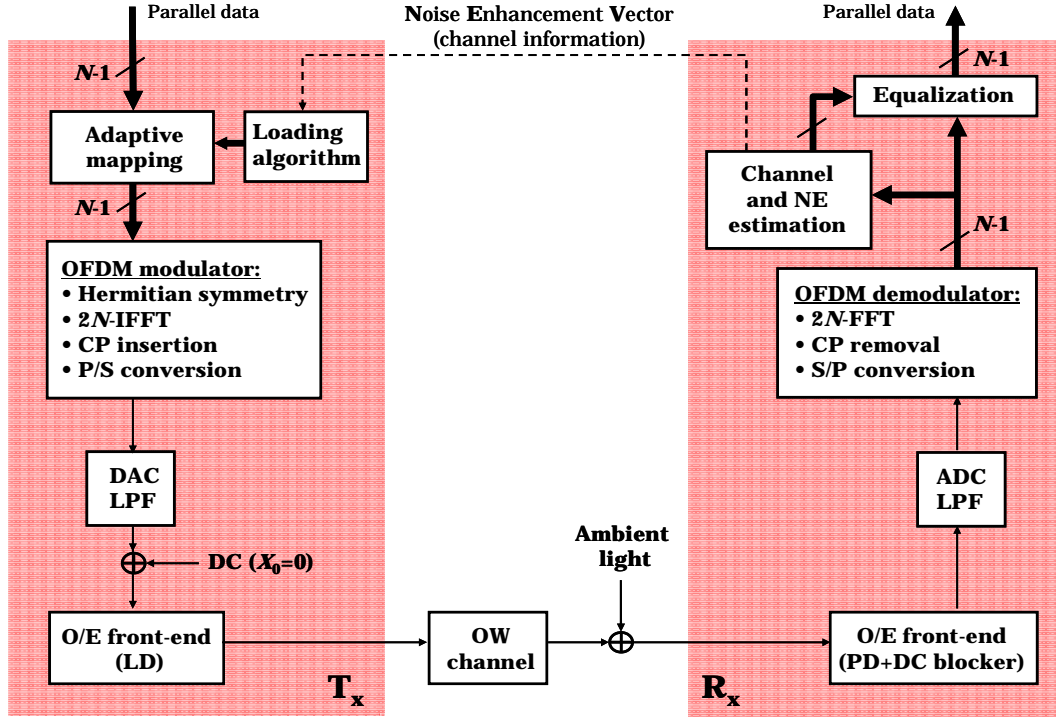


Figure 3.8: Block scheme of an adaptive OW system based on real-valued OFDM. For acronyms see the text or the list at the beginning of the document.

In contrast to conventional OFDM systems used in radio communications (which are based on coherent detection), deployment of in-phase/quadrature (I/Q) modulators in IM/DD optical wireless systems is not practical. First, due to direct detection, phase information of the optical carrier is lost. It goes without saying that transmission only of the I-component (or Q-component) would result in erroneous data after the FFT at the receiver (since the information transmitted is not complete). There is an option to use an intermediate carrier frequency, but the instability issues make the analogue I/Q modulators unattractive, especially for OFDM signals. A digital realization of I/Q modulators (at an intermediate frequency) is possible, but in order to avoid having an extra component, we adopt a different approach, where a real-valued signal is generated directly at the output of the IFFT block. For this, we use a well known DFT property. An N -DFT of a real-valued sequence has conjugated symmetric coefficients around the point $N/2$ [56]. This means that by enforcing the conjugate-symmetry of the input vector $[\mathbf{X}]$ before the IFFT, a real-valued IFFT output can be obtained. With N (actually $N - 1$) independent subcarriers foreseen in the system, a $2N$ -IFFT block

is required for a real-valued OFDM symbol. Conjugate symmetry (often referred to as *Hermitian symmetry*) on the input vector $[\mathbf{X}] = [X_0 X_1 \dots X_{N-1}]$ results in a vector of $2N$ elements defined by

$$\begin{aligned} X_{2N-n} &= X_n^*, \quad n = 1, 2, \dots, N-1, \\ X_N &= 0, \quad X_0 \in \mathbb{R}, \end{aligned} \quad (3.22)$$

which is fed into the IFFT block. The first input X_0 , corresponding to the zero frequency, needs to be real-valued and, as already mentioned, is generally left unmodulated. Obviously, the need for Hermitian symmetry at the input requires doubling the size of the IFFT block (which for a moderate number of subcarriers considered in optical wireless systems presents no implementation issue)⁸, it allows however a digital implementation of multicarrier modulation (without prohibitive analogue filter banks). The approach with real-valued IFFT output is also used in DSL systems, since the DSL channel is practically a baseband one and there is no need for I/Q modulators to shift the spectrum to a higher center frequency (as in RF applications). This version of OFDM (with real-valued IFFT output) is usually referred to as discrete multitone [46], [45].

If the input vector $[\mathbf{X}]$ has symbols of complex-valued modulation formats (usually M-QAM) as elements, so that $X_n = a_n + jb_n$, time samples at the output of $2N$ -IFFT (assuming Hermitian symmetry enforced at its input) are

$$\begin{aligned} x(k) &= \frac{1}{2N} \sum_{n=0}^{2N-1} X_n e^{j2\pi \frac{nk}{2N}} \\ &= \frac{1}{2N} X_0 + \frac{1}{2N} \sum_{n=1}^{N-1} X_n e^{j2\pi \frac{nk}{2N}} + \frac{1}{2N} \sum_{n=1}^{N-1} X_n^* e^{-j2\pi \frac{nk}{2N}} \\ &= \frac{1}{2N} X_0 + \frac{1}{2N} \sum_{n=1}^{N-1} 2\Re \left\{ X_n e^{j2\pi \frac{nk}{2N}} \right\} \\ &= \frac{1}{2N} X_0 + \frac{1}{N} \sum_{n=1}^{N-1} \left(a_n \cos(2\pi \frac{nk}{2N}) - b_n \sin(2\pi \frac{nk}{2N}) \right) \\ &= \frac{1}{2N} X_0 + \frac{1}{N} \sum_{n=1}^{N-1} \sqrt{a_n^2 + b_n^2} \cos \left(\frac{2\pi nk}{2N} + \arctan \frac{b_n}{a_n} \right), \end{aligned} \quad (3.23)$$

for $k = 0, 1, \dots, 2N-1$. In (3.23), apart from the conjugate-symmetry property of the input vector (3.22), symmetry property of the DFT twiddle factors $e^{j2\pi(N-k)k/N} = e^{-j2\pi nk/N}$ has also been exploited (see Fig. B.2 in Appendix B).

⁸The complexity of the FFT operation scales with $N \log_2 N$, according to the Cooley-Tukey algorithm [56].

At the IFFT output, a sum of $N - 1$ sampled real-valued cosinusoids is obtained. If the signal limited to B in base-band is divided among N independent subcarriers with $\Delta f = B/N$ spacing, then, because of the needed Hermitian symmetry, frequencies in IFFT cover the bandwidth of $2B$. The period of an OFDM symbol (here still without cyclic prefix), containing $2N$ time samples, is $T_{\text{FFT}} = 1/\Delta f$ (mind that time interval T_{FFT} has been previously referred to T_{sym} , but now, in order to differentiate more easily between the symbol lengths with and without cyclic prefix, we introduce new notation), leading to the sample interval $T_{\text{sam}} = 1/2B$ in time. According to the sampling theorem [53], [50] this is enough to completely determine the continuous signal at the output of the DAC. After IFFT, cyclic prefix of length $T_{\text{CP}} = LT_{\text{sam}}$ is added so that *the actual duration* of a real-valued OFDM symbol is $T_{\text{OFDM}} = T_{\text{CP}} + T_{\text{FFT}}$, and the continuous signal after DAC is

$$x(t) = \frac{A_0}{2N} + \frac{1}{N} \sum_{n=1}^{N-1} A_n \cos(2\pi f_n t + \varphi_n), \quad -T_{\text{CP}} \leq t < T_{\text{FFT}}, \quad (3.24)$$

where $A_n = \sqrt{a_n^2 + b_n^2}$ and $\varphi_n = \arctan(b_n/a_n)$ are the amplitude and the initial phase of each cosinusoid on the frequency occupied by the n^{th} subcarrier, determined by the amplitude and phase of the complex-valued symbol X_n at the corresponding IFFT input. The equality

$$2\pi \frac{nk}{2N} = 2\pi n \frac{2B}{2N} \frac{k}{2B} = 2\pi n \Delta f k T_{\text{sam}} = 2\pi f_n t_k \quad (3.25)$$

is also used, where $f_n = n\Delta f$, $n = 1, 2, \dots, N-1$ and $t_k = kT_{\text{sam}}$, $k = 0, 1, \dots, 2N-1$. Note that the signal $x(t)$ would also be obtained with a bank of $N - 1$ analog filters. After DAC, an analogue low-pass filter (LPF) is implemented to suppress the aliasing spectra. At the same time, this filter performs interpolation of the discrete signal waveform.

Figure 3.9 illustrates an example of the signals at the input and output of the real-valued OFDM modulator. In this example, 3 out of $N = 8$ orthogonal carriers in $B = 20$ MHz are modulated by 16-QAM to generate an input vector (before symmetry enforcement) $[\mathbf{X}] = [0, 1 + j1, 3 - j1, 0, -3 + j3, 0, 0, 0]$. The input of 16-IFFT is shown in Fig. 3.9 a, the output is shown in Fig. 3.9 b, with separate contributions of individual subcarriers given in Fig. 3.9 c. (The output levels in the figure are scaled by N .)

Another difference to RF systems, in general needed for multicarrier optical wireless transmission, is a positive bias. The modulating signal (AC part of $x(t)$) contains also negative values, thus a DC component is required to properly drive the optical source (since the optical power can only be positive). The bias can be introduced directly at the system input by setting appropriate X_0 , or, if $X_0 = 0$ is left, DC can be added directly in front of the optical source. The bias is, however,

limited by safety regulations, and together with the requirement for non-negative modulating signals, it constrains the achievable transmission rate. This issue is thoroughly addressed throughout Chapters 5 and 6.

Returning to Fig. 3.8, the signal reaches the receiver after undergoing the influences of time-dispersive channel and ambient light (dominant source of noise in the channel, introduced in Section 2.4), where the bias signal is discarded by the AC-coupled detector. After detection, digitalization by an analogue-to-digital converter (ADC), and OFDM demodulation (i.e., cyclic prefix removal and FFT), thanks to the preserved orthogonality, subcarriers are separately processed. In principle, only the outputs $n = 1, \dots, N - 1$ need to be further regarded, since they are the ones carrying information. The data are equalized, i.e., multiplied with the inverse channel coefficients, which are estimated using a training sequence. The required period of channel information updates is addressed in Chapter 6. After estimation of channel coefficients, noise enhancement on each subcarrier (power of the second summand in (3.21)) is acquired, and this noise enhancement vector (NEV) is transmitted over the feed-back link. Noise enhancement can be interpreted as the noise influence translated to the transmitter side, and therefore presents information needed for the loading algorithm. Note that the required feedback link can be easily made reliable in OW channel if it is set to occupy the lowest subcarrier(s) of the uplink, which are always suitable for transmission (see Fig. 2.10) thanks to the diffuse signal component. This favorable property is not the case in the radio channel, where fading holes are likely to occur anywhere in the used bandwidth and diversity-oriented schemes would be used instead.

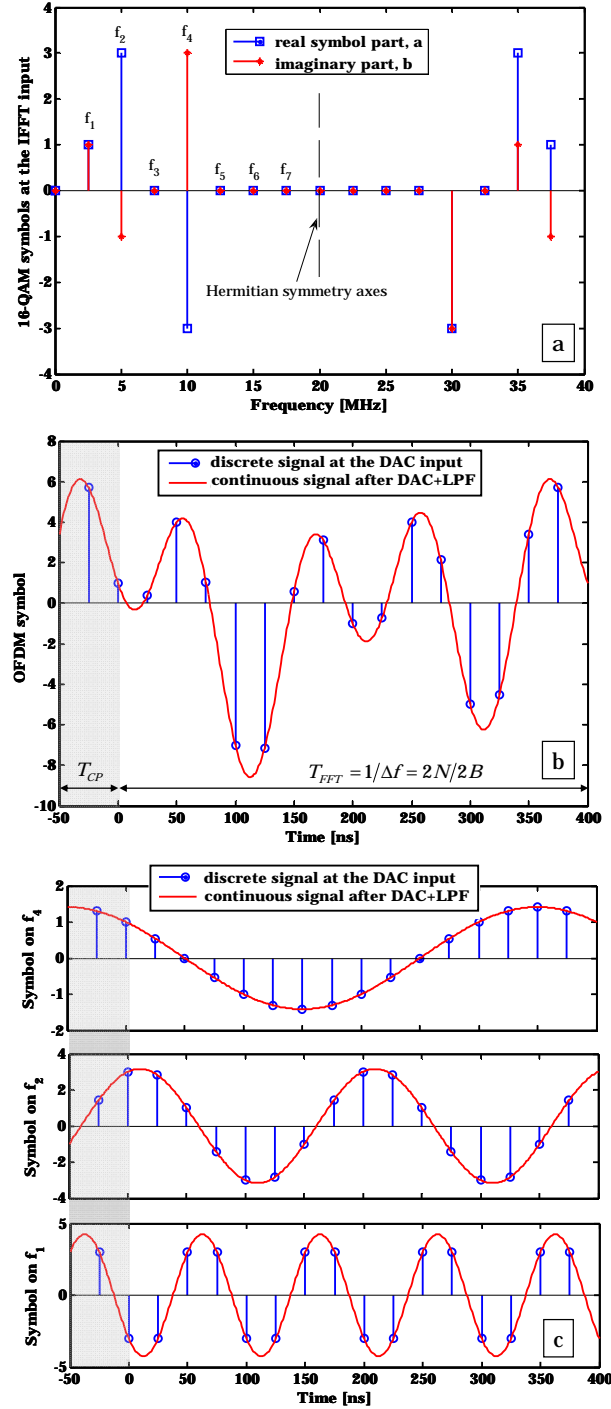


Figure 3.9: Example of the signals at the input and output of IFFT, when $N = 8$, $B = 20$ MHz, and $L = 2$. Panel a) shows 16-QAM symbols on individual subcarriers in f -domain, panel b) shows the OFDM symbol in t -domain, and c) the individual contributions of modulated subcarriers. For t -domain both sampled and continuous signals before and after the DAC are shown, scaled by N to achieve the same level as in f -domain.

Chapter 4

Design of Simple Opto-Electrical Front-Ends

As a prerequisite for investigation of the transmission technique in focus, Chapter 4 analyzes and defines the opto-electrical front-ends of the system. The aim is to keep the component design simple and to allow broadband transmission with enough link budget at the same time. In that sense, in Section 4.1.1, a maximum permissible transmitter power is obtained in accordance with safety regulations. In Section 4.1.2, the number of subcarriers is discussed in the course of the source linearity analysis. On the other side of the OW link, components of a simple infrared receiver are discussed and dimensioned (so that enough bandwidth for a broadband signal is provided), followed by a noise analysis, all in Section 4.2. The system parameters obtained in this way are then used to support the numerical examples in the subsequent chapters.

4.1 Optical source

The two most popular solid-state light emitting devices are *light emitting diodes* (LEDs) and *laser diodes* (LDs), operating in forward bias. As a consequence of a drive current at the device input, a large portion of injected carriers recombine and generate photons. The recombination process in the active region of an LED occurs randomly, and the photons generated by such *spontaneous emission* are incoherent (i.e., they have a random phase). On the other hand, in LDs the majority of generated photons are coherent (i.e., same phase) resulting from the *stimulated emission*. Also, while LEDs emit an optical intensity approximately linear to the drive current, lasers are threshold devices - at low drive currents, spontaneous emission dominates and the device behaves like a low intensity LED, however, after the current exceeds the threshold level, stimulated emission dominates which leads to a high optical efficiency of the device.

Table 4.1 presents a comparison of LED and LD characteristics, relevant for optical wireless links. Even though LDs are more complex and more expensive than LEDs, they possess several crucial assets. The main advantage of LDs is in the speed of operation. Due to the stimulated emission, LDs can operate at pulse rates in the GHz range, whereas modulation bandwidths of typical low-cost LEDs are limited to tens of MHz. Wider LED modulation bandwidths are usually obtained at the cost of reduced electro-optic conversion efficiency. Another desirable quality of laser diodes is their very narrow spectral width, which allows the use of narrow optical pass-band filter at the receiver, resulting in superior rejection of background radiation. The stimulated emission also leads to much higher optical efficiencies of LD than of LED devices. On the other hand, the property which can be seen as the main advantage of LEDs over LDs in OW links is their safety. Even though no difference is made between LEDs and LDs in the safety standards, because they emit light from a sufficiently large surface area and with wide emission angles, most LEDs are generally considered eye safe. As shall be seen in more detail in the Section 4.1.1, the lasers are considered as point sources and need to be rendered eye safe with help of a *diffusor*. A diffusor is an element which destroys the spatial coherence of the laser output and spreads the radiation over an extended emission aperture and angle. An example of a typically used diffusor is a thin piece of transparent plastic with rough surface, which generates a Lambert radiation pattern and has an efficiency of about 70 %. Another simple way to obtain a Lambertian diffusor is to use an integrating sphere in front of the laser. A more sophisticated type are the computer generated holograms, which allow free tailoring of the radiation pattern and efficiencies of up to 100 %, but they are quite complex to fabricate [67].

Because of the mentioned advantages, the optical sources for the system considered in this work are assumed to be laser diodes, in combination with a simple transmissive diffusor.

4.1.1 Safety considerations

Human beings can see wavelengths between 400 and 700 nm. This is most probably the result of evolution since this range corresponds to the peak of the solar

Characteristic	Light Emitting Diode	Laser Diode
Modulation bandwidth	up to ≈ 100 MHz	up to tens of GHz
Spectral width	25 to 100 nm	10^{-5} to 5 nm
E/O efficiency	10 to 40 %	30 to 70 %
Safety	inherently safe	needs a diffusor
Complexity and cost	low	moderate

Table 4.1: Summary of LED and LD characteristics relevant for OW links based on [51].

spectrum. Nevertheless, specific tissues in the eye interact strongly also with other portions of optical radiation. As all tissues, tissues of eye are susceptible to damage under severe conditions, such as when the eye is in the path of radiation from a high-intensity source.

Various parts of the eye are differently susceptible to damage at a given wavelength. The *cornea* (glassy front of the eye) and the *lens* (behind the cornea and in front of the retina) focus the visible light onto the *retina* (the back surface of the eyeball that detects light). Cornea and lens are transparent to the visible and near-IR wavelengths (up to 1400 nm). These wavelengths are therefore quite dangerous to the retina, as the eye is intended to be sensitive here. The lack of natural aversion reflex (so-called “blink, look away”), makes the near-IR radiation particularly hazardous when subjected to strong radiation. Among several mechanisms by which an injury can occur, the retina is most vulnerable to excessive thermal interaction which can cause permanent damage of vision due to the retinal burn. On the other hand, mid and far infrared radiation are absorbed by cornea and lens (nothing passes through to the retina). In these regions, damage of lens and cornea may occur, but constraints are less critical. In fact, compared to the human skin, which is the other tissue very susceptible to damage, beyond 1400 nm, the danger for the two tissues (eye and skin) is the same. The biological effect of irradiation of skin by lasers operating in the visible and infra-red spectral regions may vary from a mild erythema to severe blisters. Eye and skin safety are both regarded in this thesis.

Lasers are potentially dangerous as a result of their high *radiance* (expressed in $[W/m^2sr]$).¹ For example, a laser has a radiance many times higher even than direct sunlight. Admittedly, the sun has a much larger total output power, but its radiation is spread over a large surface and a large solid angle (4π). The laser produces a beam of light over a solid angle several thousand times smaller, resulting thus in much higher radiance. The collimated radiation (a “parallel” beam of radiation with very small angular divergence or convergence) of a laser renders also a high power density, called *irradiance* ($[W/m^2]$), because it can be focused to a much smaller spot on the retina than conventional light sources.

There are a few international standard bodies which provide safety guidelines on laser transmission, such as International Electrotechnical Commission (IEC), American National Standards Institute (ANSI), and European Committee for Electrotechnical Standardization (CENELEC). The IEC standard IEC 60825 – 1 [68] is widely accepted and is considered here.

¹The radiance (sometimes also referred to as brightness) of an optical source is defined as the source power P per unit solid angle Ω (source intensity), divided by the source area projected in the observation direction $P/(\Omega A \cos \theta)$.

The norm defines the *maximum permissible exposure* (MPE) as a measure of safety of an optical source. This is the level of electromagnetic radiation to which a human observer may be exposed over some time without incurring adverse effects immediately or after a longer period. The values specified by the norm are set an order of magnitude lower than the experimentally obtained hazardous levels. MPE depends on **1)** the wavelength of the radiation, **2)** the mode of source operation (in the case of continuous wave (CW) operation,² MPE depends on the exposure time and is expressed by power density in $[\text{W}/\text{m}^2]$, whereas in the case of pulsed or modulated source, MPE regards the pulse duration and is expressed by the density of energy in $[\text{J}/\text{m}^2]$), **3)** the tissue at risk and **4)** size of the image on the retina (for visible and near infrared radiation in the range of 400 – 1400 nm).

Apart from the limits on maximum permissible exposure, for manufacturers of optical sources, the norm also provides a classification scheme which defines the product emission limits for each of the classes. The classification scheme is set up independently of the laser application and involves many worst case assumptions, such as long exposure durations, close distance to the laser aperture and the use of optical instruments, all of which may increase the hazard. These levels are expressed by an *accessible emission limit* (AEL), as the maximum accessible emission level permitted within a particular class. The class indicates how safe a product is and what type of precaution measures are needed.

In optical wireless systems, where the radiation is unguided, the only practicable laser class is class 1. By definition, products pertaining to this class pose *no risk to eyes or skin* under normal operations and conditions, including occasions when users view the beam directly with optics that could concentrate the output into the eye. The accessible emission limits of Class 1 are simply derived by multiplication of the respective maximum permissible exposure values with the area of the eye pupil (7 mm in diameter, corresponding to the worst case of a completely dilated pupil [68]).

The norm provides a table by which the AEL value for a given source should be calculated, depending directly or indirectly on the factors mentioned above. Hereby, for wavelengths $400 \leq \lambda \leq 1400$ nm, a difference is made between *point* and *extended sources*, according to the angular subtense (also referred to as field-of-view) of an apparent source, when it is located in the eye's range of accommodation (≥ 100 mm). The detailed description of the measurement setup geometry and derivation of the AEL (i.e., MPE) values is given in the Appendix C.2.

²In the norm, a laser operating with a continuous output for a period equal to or greater than 0.25 s is regarded as a continuous wave laser. Sometimes it is necessary to consider a modulated laser, however, for rates higher than about 1 Mbit/s, the CW operation poses a more restrictive limit.

Figure 4.1 shows the final results of the calculations in terms of the maximum permissible source power, depending on the aperture diameter a of the apparent source. In Fig. 4.1 a, class 1 sources operating in CW regime at 900 nm and 1550 nm are considered, having either a collimated or a Lambertian characteristic with radiation index $m = 1$ and $m = 2$. Figure 4.1 b shows the maximum permissible source powers with no adverse effects on human skin. Even though results in Fig. 4.1 a are based on calculation of AEL values for the eye, and results in Fig. 4.1 b on calculation of MPE values for the skin, they are comparable, because only sources of class 1 are considered.

In all calculations, an exposure time of $t = 3 \cdot 10^4$ s is considered, to include the possibility of intentional viewing (e.g., a curious baby playing with an end-terminal device).³

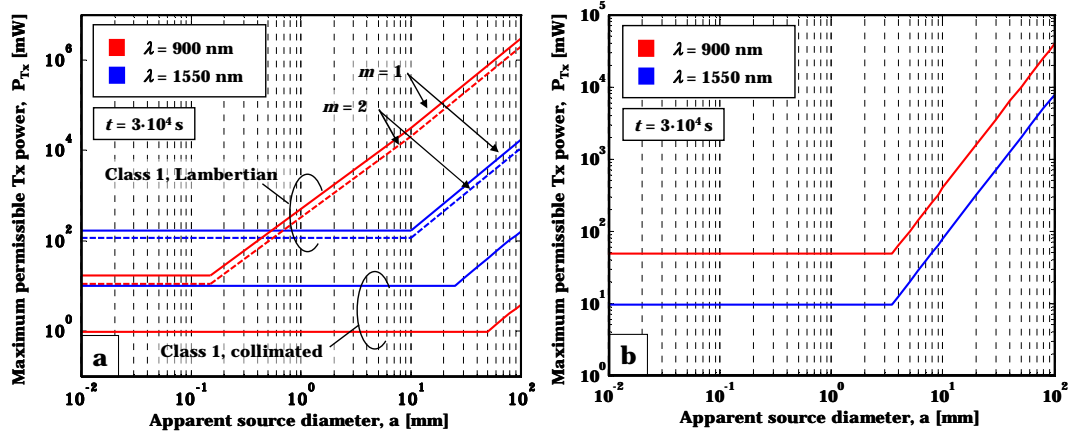


Figure 4.1: (a) Maximum permissible optical Tx power as a function of the apparent source diameter in the case of class 1 collimated beams and class 1 divergent beams with Lambertian characteristic ($m = 1, 2$). Values are based on calculation of the accessible emission limit when risk of thermal retinal injury is considered. Apparent source diameter is related to the angular subtense over the distance of 100 mm. (b) Maximum permissible optical Tx power as a function of the apparent source diameter. Values are based on maximum permissible exposure calculation when risk of skin injury is considered. Apparent source diameter is equal to the actual source diameter.

For a collimated beam, a maximum Tx power is in principle independent on the size of the apparent source, since there is practically no beam divergence (in the case of an ideally collimated beam). Higher power levels are permissible at long wavelengths, since the light in this region does not get focused onto the retina. The kinks in the graphs are a result of the testing conditions geometry (which are,

³According to the norm, for Class 1 and 1M for wavelengths longer than 400 nm, the time base is either 100 s or 30000 s, depending on whether the product is designed for long term intentional viewing (30000 s) or not (100 s).

incidently, different for the two wavelengths regarded).

In the case of a source with a highly divergent beam (i.e., with a Lambertian characteristic) radiating in the near-IR range (@900 nm), there is a clear increase in the maximum permissible power when the source is *extended* (corresponds to $a \geq 0.15$ mm). The reason is that for the divergent sources which are smaller than a certain threshold, the retinal image is assumed to be a diffraction-limited spot, regardless of the source size. Such sources are then regarded as point sources. If the apparent source size exceeds the mentioned threshold (i.e., the source is extended), the image on the retina is larger and the injury threshold in terms of the total permissible source power increases. The second kink ($a = 10$ mm) is a consequence of the testing geometry. According to the norm, there is a limiting angle of acceptance for wavelengths $302.5 \text{ nm} \leq \lambda \leq 4000 \text{ nm}$, beyond which there is no spot-sized dependence of the retinal thermal injury threshold and the permissible power increases with $\sim a^2$.

In the long-wavelength region, the risk of retinal thermal injury remains constant irrespective of the source size (since there is no focusing on the retina) and only the second kink is present. In Fig. 4.1 a, it is observable that, in case of the point sources, maximum permissible Tx power is higher at 1550 nm. However, when the source is extended, the permissible power becomes much higher for the same source type operating at 900 nm. Moreover, from the figure it is clear that less divergent (more directed) sources (case $m = 2$) are allowed to emit less power.

According to the norm, in the near-IR wavelength range human skin tolerates higher radiation levels than the eye (because of light focusing onto the retina), whereas, above 1400 nm, tissue tolerance is more or less the same for both skin and eye (compare Fig. 4.1 a and b). The kink in the characteristic is, again, a result of the measurement geometry. After the threshold, the power permitted grows with $\sim a^2$.

By comparing the two figures, it can be concluded that the highest power is permissible for an extended Lambertian source at 900 nm, but that the more critical tissue is the skin. For example, for $a = 10$ mm the allowed power level is about 400 mW. This is because the specified maximum permissible exposure level for the skin must be measured at the position where the exposure level is the highest, i.e., when the skin is in direct contact with the emitter output window. Therefore, for further calculations, a value of 400 mW is taken as the limit at Tx, while keeping in mind that this limit may be lifted as long as direct contact is prevented (i.e., a minimum size of a fully illuminated surface is enforced).⁴

⁴Recently, a new version of the standard was issued [69] (however, a German norm is yet pending), where the conditions regarding eye safety are even more relaxed. Due to the fact that the work has already been conducted following [68] and that skin safety considerations remained

4.1.2 Transmission of analogue multitone signals

Multitone signals in nonlinear systems

An OFDM-based system uses several low-rate sub-carriers to transmit data and is efficient in combating the effects of time-dispersive channels. It is, at the same time, very sensitive to the nonlinearity of the transmission system. An important phenomenon, responsible for this, is a signal distortion due to generation of higher-order products by the nonlinear system components.⁵

With N_F tones occupying the frequencies f_1, \dots, f_{N_F} at the input of a nonlinear system, the output signal can be composed of all the frequencies f_{out} presenting a linear combination of the input tones

$$f_{\text{out}} = s_1 f_1 + s_2 f_2 + \dots + s_{N_F} f_{N_F}. \quad (4.1)$$

The coefficients $s_n \in \{0, \pm 1, \dots\}$ define the *order* of the output product as

$$\mathcal{O} = \sum_{n=1}^{N_F} |s_n|. \quad (4.2)$$

Hence, apart from the fundamental tones, the output signal consists of many undesired components which can cause signal distortion. All distortion products have an order $\mathcal{O} \geq 2$. By its very presence at the input, each tone f_n generates at the output a *harmonic distortion* (HD) product of the k^{th} order at $k f_n$. Moreover, in interaction with the other tones, it generates *intermodulation distortion* products (IMP) whose frequencies and orders are determined according to (4.1) and (4.2), respectively. For example, a tone at frequency f_1 together with a tone at f_2 leads to a third order IMD at $2f_1 - f_2$.

When distortion products fall within the transmission signal pass-band, filtering them out becomes practically impossible and, in such cases, they cause the most damage. Especially for equally spaced tones ($f_n = n f_1$) of interest in this thesis, the in-band products fall directly onto the fundamental frequencies. Moreover, several different harmonic or intermodulation products may accumulate at the same frequency (e.g., $f_3 - 2f_1 = f_5 - 2f_2 = f_1$) creating particularly strong distortions.

Nonlinearity in analogue OW links

Nonlinearity is of importance for all systems where transmission of analogue, especially multitone, signals is considered; not only optical wireless, but also conventional fiber-optics, radio-over-fiber, hybrid radio-coax systems, etc. Severity

unchanged in [69] the work was not repeated.

⁵Another factor is a high peak-to-average-power ratio which, thanks to system nonlinearity, leads to signal distortion. The influence of high peaks is discussed in more detail in Chapter 6.

of distortion effects can vary among different systems, due to different nonlinear components and modulating frequencies.

The regarded optical wireless system considers a laser diode directly modulated by analogue signals at relatively low frequencies (up to 100 MHz). In OW links, the nonlinear response of LD presents the main source of signal degradation by nonlinear distortions.⁶ Moreover, at the assumed modulation frequencies, the imperfect linearity of the static optical-power-vs-current laser characteristic (P - I) can be held responsible for this effect [70].

In general, the P - I characteristic can be described by a Taylor expansion around the operating point $(\langle I \rangle, \langle P \rangle)$

$$P(I) = \langle P \rangle + \sum_{l=1}^{\infty} \frac{1}{l!} \left. \frac{\partial^l P}{\partial I^l} \right|_{(\langle I \rangle, \langle P \rangle)} (I - \langle I \rangle)^l. \quad (4.3)$$

The sum of the first k summands ($l = 0, 1, \dots, k-1$) includes distortions up to the order $k-1$. Usually, an idealized P - I characteristic is considered, where the summands $l \geq 2$, and the nonlinear effects with them, are neglected. In this subsection, however, the influence of second and third order distortions is characterized and evaluated for the transmission of OFDM signals. For such consideration, Taylor expansion (4.3) can be limited to the first four summands ($l = 0, \dots, 3$)

$$P(I) = \langle P \rangle + S(I - \langle I \rangle) + \frac{1}{2}S_2(I - \langle I \rangle)^2 + \frac{1}{6}S_3(I - \langle I \rangle)^3, \quad (4.4)$$

where S , S_2 and S_3 denote the first, second and third derivative of the characteristic in the operating point. Also, it is sufficient to assume an input signal $I(t)$ consisting of three equidistant tones at $f_n = nf_1$, $n = 1, 2, 3$ modulated with the same modulation depth m (i.e., which have the same amplitude $A_n = m_n \langle I \rangle$)

$$I(t) = \langle I \rangle + m_n \langle I \rangle \cos(2\pi f_1 t) + m_n \langle I \rangle \cos(2\pi f_2 t) + m_n \langle I \rangle \cos(2\pi f_3 t). \quad (4.5)$$

After introducing (4.5) in (4.4) and performing simple manipulations, the output signal can be arranged to reveal how each tone generates a second and third harmonic (HD2 and HD3), each pair of tones generates a second order intermodulation product (IMP2) at $f_n \pm f_j$, and a third order intermodulation product (IMP3) at $2f_n \pm f_j$, while each triplet generates an IMP3 at frequencies $f_n \pm f_j \pm f_k$. Also, on account of trigonometric equalities

$$\begin{aligned} \cos^2(x) &= \frac{1}{2} + \frac{1}{2} \cos(2x) \quad \text{and} \\ \cos^3(x) &= \frac{3}{4} \cos(x) + \frac{1}{4} \cos(3x), \end{aligned} \quad (4.6)$$

⁶In a conventional fibre-optic link, nonlinearities of the external modulator (when present) and fibre infrastructure often contribute to much higher signal degradations.

each tone contributes to an increase of the DC signal and to an increase of the signal amplitude on fundamental frequency (F in the table), meaning that there is a weak amplification of the desired signal. Even though DC contributions do not cause in-band damage, they lead to the increase of the mean optical power, which needs to be kept within the safety limits. The significance of this effect is evaluated towards the end of this subsection. The overview of the tones generated at the laser output, together with distortion type, order and amplitude, is given in the table in the left part of Fig. 4.2, whereas the bar graph illustrates the distribution of distortion products and weight of their influence on the frequency grid.

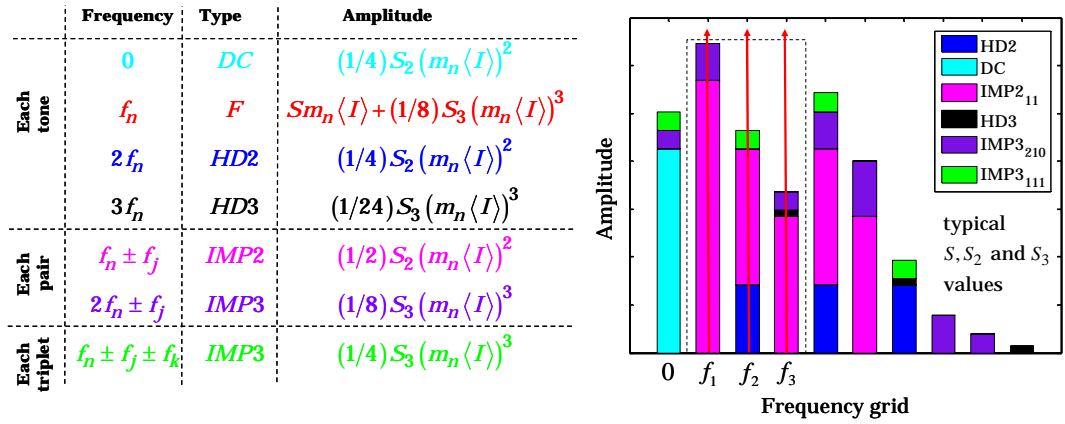


Figure 4.2: Table: Overview of the tones generated at the laser output together with distortion type, order and amplitude. Bar graph: Distribution of second and third order harmonic and intermodulation products on the frequency grid, with three equally spaced equally modulated fundamental tones at the laser input. The dashed rectangle denotes the pass-band of the input signal.

It is clear that the distortion products have different amplitudes which mainly depend on their order. Namely, the second (third) order distortions have amplitudes proportional to the square (cube) of the fundamental tone amplitudes. Figure 4.2 shows the three fundamental tones represented by arrows and all the distortion products as bars of different colors, matching the colors in the table. The amplitudes in the bar graph are calculated for typical values of S , S_2 and S_3 coefficients,⁷ and normalized by the fundamental tone amplitude. Because the output amplitudes of pure fundamental tones are much larger than the amplitudes of the other tones, arrows are used to indicate that they are not presented to scale.

From the figure, it can be concluded that even with only 3 tones at the input, a multitude of output components appears. Several distortions of one or different types superimpose on the same fundamental frequency. Obviously, if the number

⁷The values for the coefficients are obtained by measurements of one typical vertical-cavity surface-emitting laser (VCSEL). This is presented later in the text.

of input tones with the same amplitude increases, the significance of the nonlinearities grows, and can seriously determine the system performance. From Fig. 4.2, it can also be anticipated that the IMP2 is the distortion type of the largest impact (as will be confirmed by results that follow), but theoretically this depends on the S_2 and S_3 coefficients.

Qualitative limits on the modulating signal

Let the modulating signal (without DC) at the laser input be a sum of N_F signals on equidistant subcarrier frequencies

$$i(t) = \sum_{n=1}^{N_F} A_n \cos(\varphi_n + 2\pi f_n t), \quad (4.7)$$

where A_n and φ_n denote the amplitude and initial phase of the signal on $f_n = n f_1$. In principle, this corresponds to the OFDM modulating signal defined by (3.24) (here scaled by N , without loss of generality). Note that the number of tones at the laser input N_F in (4.7) is equal to $N - 1$ in (3.24).

For the sake of simplicity let all tones be modulated with the same modulation index m_n , so that $A_n = m_n \langle I \rangle = \text{const.}$ Since φ_n is random, for more than just a few tones at the input, a Gaussian distribution of the signal in front of the laser can be assumed, with a mean value $\langle I \rangle$ and a standard deviation $m_{\text{eff}} \langle I \rangle$, where

$$m_{\text{eff}} = \sqrt{\sum_{n=1}^{N_F} m_n^2} = m_n \sqrt{N_F}. \quad (4.8)$$

Figure 4.3 illustrates the P - I characteristics of an idealized (Fig. 4.3 a) and realistic (Fig. 4.3 b) laser together with a modulating signal. An idealized optical source has a threshold at 0 and it is perfectly linear. System linearity can be conveniently quantified with a *carrier-to-interference* (C/I) ratio, defined as the ratio of electrical powers of the desired output signal and total interference at the signal frequency. In the case of the idealized optical source, C/I is infinitely large. Hence, from the linearity point of view, the laser operating point ($\langle I \rangle, \langle P \rangle$) could be set arbitrary high. The only limit on the signal (and, thus, on the achievable rate) is then given by the safety constraint and the required signal non-negativity or error performance if clipping is permitted in front of the optical source. While the former limits the mean optical power to $P_O = \langle P \rangle$ as $\langle I \rangle \leq P_O/S$, the latter limits the effective modulation depth of the signal.

On the other hand, the imperfect linearity of the realistic P - I characteristic influences the choice of the laser operating point. Around the operating point, C/I is finite, and for a given number of tones, it decreases when the modulation index

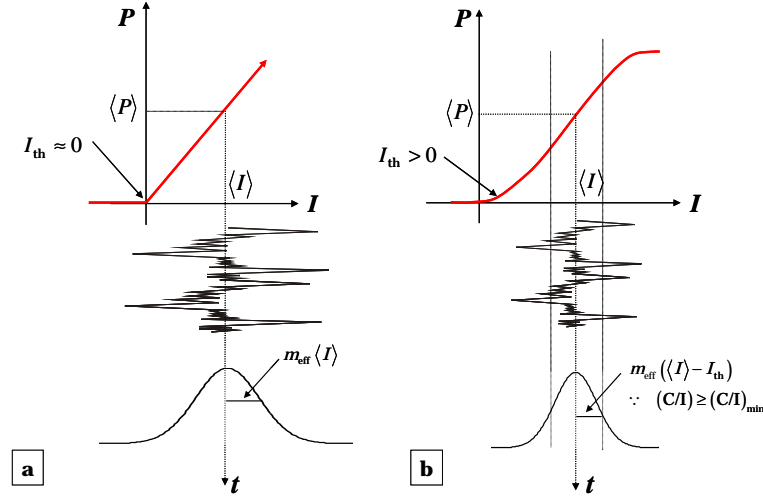


Figure 4.3: Sketches of a) idealized (perfectly linear) and b) realistic static P - I laser characteristic.

m_n is increased. For a chosen operating point, the acceptable C/I ratio defines the maximum value of m_{eff} for which the modulating signal doesn't suffer significant consequences of the nonlinear laser characteristic. In short, the limit on the effective modulation index with N_F tones at the input can be described as

$$m_{eff}(N_F) \leq m_{eff}^{max}(N_F), \quad m_{eff}^{max}(N_F) \text{ so that } (C/I) \geq (C/I)_{min}. \quad (4.9)$$

Equivalently, for a certain m_{eff} (certain modulating signal power), the minimum permissible C/I ratio limits the number of fundamental tones at the laser input, i.e., for a given $(C/I)_{min}$, there is a trade-off between m_{eff} and N_F . Moreover, in the case of a realistic laser characteristic, the threshold current I_{th} can not be approximated with 0, contributing further to the limit of m_{eff} . The goal of the following investigations is, therefore, to determine how nonlinearities influence the C/I ratio when a relatively large number of tones is considered at the input of a typical commercially available optical source component, which could be deployed in the considered system.

Experimental investigation of a VCSEL linearity region

In order to obtain a fair model of the laser characteristic (4.4) and investigate the influence of nonlinearities with N_F tones at the laser input, a typical 840 nm vertical-cavity surface-emitting laser (VCSEL) is experimentally characterized (i.e., a suitable operating point is chosen, and coefficients S , S_2 and S_3 evaluated).

In the first step, the static P - I characteristic of the VCSEL is measured with the setup illustrated in Fig. 4.4 a and shown in Fig. 4.4 b. The laser is driven by a constant current source over a bias tee, and the optical power is directly captured by

the sensor. Immediately, the laser threshold current $I_{th} = 17$ mA can be evaluated.

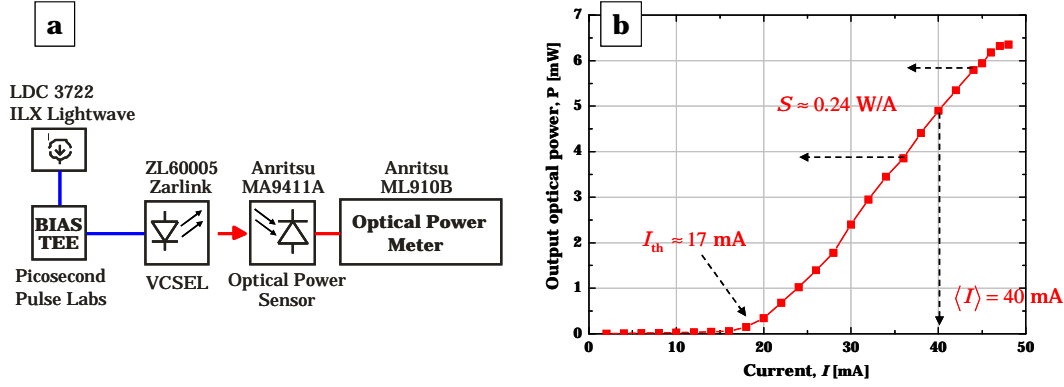


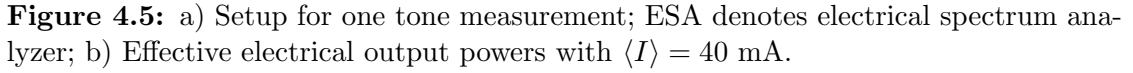
Figure 4.4: Static P - I characteristic measurement a) setup b) result.

In order to choose a suitable operating point, additional measurements are performed with the setup depicted in Fig. 4.5 a. A sinusoid is generated by a synthesizer of constant power, and its amplitude is varied by an attenuator. After the low-pass filters, a relatively pure sinusoidal tone is obtained in front of the laser, with minimum 65 dB difference between the fundamental and its harmonics.⁸ The DC signal generated by the constant current source determines the operating point $(\langle I \rangle, \langle P \rangle)$. The light from the directly modulated VCSEL is detected by a silicon p-doped-intrinsic-n-doped photodiode (Si-PIN), and the received electrical power $P_{el,out}$ is read at the electrical spectrum analyzer.⁹

Further on, a single tone (at 51 MHz and with power $P_{el,in}$) is generated at the laser input. For different levels of the input power, the powers of the fundamental tone and harmonics ($P_{el,out}$) are evaluated at the photodiode. After repeating the procedure for several operating points, suppressions of the harmonics are compared for the same input power levels. For further investigations $\langle I \rangle = 40$ mA is chosen since it allows good harmonic suppression without driving the VCSEL into saturation. The operating point could be set so high because in the dynamic regime (the case of “one tone” measurements, Fig. 4.5), the P - I characteristic is more linear towards the high-end (compared to the static characteristic shown in Fig. 4.4 b, obtained with CW operation), since the influence of the thermal effects is not significant. This was verified by performing some measurements of the P - I characteristic with the laser operated in pulse-mode.

⁸If the tones at the laser input have significant harmonics, these behave as independent tones when interaction at the laser takes place. However, the achieved suppression of 65 dB allows to disregard their influence.

⁹The photodiode was terminated by a 50 Ω resistor, so the electrical spectrum analyzer showed actually only the half of the received power. $P_{el,out}$ represents the correct received power.


$$S \approx \Delta P / \Delta I \approx 0.24 \text{ W/A}. \quad (4.10)$$
$$P_{\text{el, in}} [\text{dBm}] = 10 \log \left(\frac{Z_{\text{LD}}}{2} \frac{A_n^2}{1 [\text{mW}]} \right), \quad (4.11)$$
$$\begin{aligned} P_{\text{el,out}}(\text{F}) \text{ [dBm]} &= 10 \log \left(\frac{Z_{\text{PD}}}{2} \frac{(H_{\text{LD-PD}})^2}{1 \text{ [mW]}} \left(S A_n + \frac{1}{8} S_3 A_n^3 \right)^2 \right), \\ P_{\text{el,out}}(\text{HD2}) \text{ [dBm]} &= 10 \log \left(\frac{Z_{\text{PD}}}{2} \frac{(H_{\text{LD-PD}})^2}{1 \text{ [mW]}} \left(\frac{1}{4} S_2 A_n^2 \right)^2 \right), \\ P_{\text{el,out}}(\text{HD3}) \text{ [dBm]} &= 10 \log \left(\frac{Z_{\text{PD}}}{2} \frac{(H_{\text{LD-PD}})^2}{1 \text{ [mW]}} \left(\frac{1}{24} S_3 A_n^3 \right)^2 \right), \end{aligned} \quad (4.12)$$

where Z_{PD} and η denote serial resistance and responsivity of the photodiode, respectively, and H_{LD-PD} represents the total attenuation between the VCSEL input and the photodiode output.

From Fig. 4.5 b, it is observable that up to a certain power ($P_{el,in} \approx 4.6$ dBm), harmonics behave as predicted by theory. Namely, according to (4.12), electrical powers at the output relate as $P_{el,out}(F)/P_{el,in} \sim 1$,¹⁰ $P_{el,out}(HD2)/P_{el,in} \sim m^2$ and $P_{el,out}(HD3)/P_{el,in} \sim m^3$, corresponding to the slopes of 1:1, 2:1 and 3:1, respectively. A slope 3:1, for instance, means that the logarithm of the third harmonic component grows 3 times faster than the fundamental tone. Therefore, for relatively low input powers HD2 is stronger than HD3, but the power of HD3 increases faster. For $P_{el,in} > 4.6$ dBm, the behavior changes, because the proximity of the threshold introduces nonlinear distortions of higher order.

From the region where the behavior of the harmonics follows the theoretical predictions, expressions (4.12) can be used to evaluate the remaining coefficients S_2 and S_3 . For this evaluation, the factor in $P_{el,out}(F)$ dependent on A_n^3 is neglected. With S obtained in (4.10), we find

$$\begin{aligned} S_2 \text{ [W/A}^2\text{]} &= 4 \frac{S}{A_n} 10^{-(P_{el,out}(F) - P_{el,out}(HD2))[\text{dB}]/20}, \\ S_3 \text{ [W/A}^3\text{]} &= 24 \frac{S}{A_n^2} 10^{-(P_{el,out}(F) - P_{el,out}(HD3))[\text{dB}]/20}. \end{aligned} \quad (4.13)$$

Calculating by (4.13) and averaging over all available measured points renders the values $S_2 = 2.7 \text{ W/A}^2$ and $S_3 = 300 \text{ W/A}^3$.

Interference with N_F tones at the laser input

Previously elaborated measurements provided the parameters $\langle I \rangle$ (and I_{th}), S , S_2 and S_3 , necessary for modeling of the laser P - I characteristic (4.4). The model is implemented in simulations to evaluate the influence of nonlinearities in the general case of N_F tones present at the source input.

As already indicated, N_F equidistant tones are assumed with equal modulation indices $m_n = m_{eff}/\sqrt{N_F}$. In the following investigation, m_{eff} (power of the modulating signal) is kept constant, while the number of tones is varied. In this way, the distribution of the modulating signal remains the same around the operating point, but the modulation depth of each tone varies with the number of tones (see (4.8)). Figure 4.6 shows the impact of the strongest distortion type (second order intermodulation products at $f_n \pm f_j$, denoted as IMP2₁₁ in Fig. 4.2) across the frequency grid, for different values of N_F and for two m_{eff} parameters.

¹⁰The second summand of the $P_{el,out}(F)$ logarithm argument is neglected.

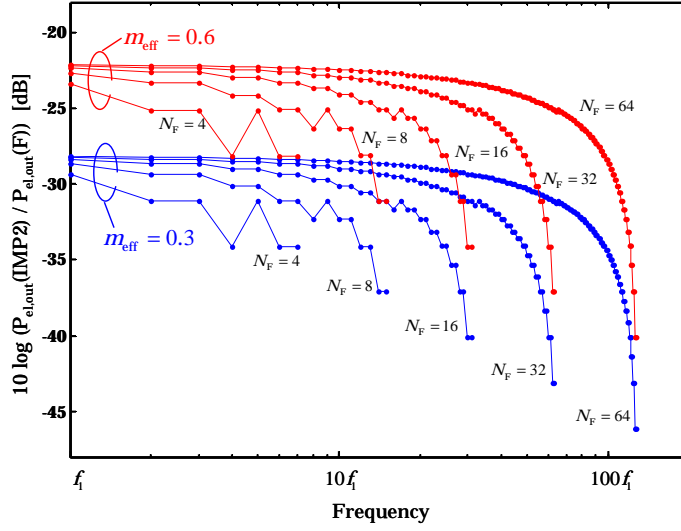


Figure 4.6: Simulated impact of the second order intermodulation distortions across the frequency grid for different numbers of tones at the input, in case when $m_{\text{eff}} = 0.6$ and $m_{\text{eff}} = 0.3$.

The y -axis shows the electrical output power of the distortion, normalized by the output power of the fundamental tone. From the figure, it is observable that the distortions induce the most damage at the lowest fundamental frequencies (in particular f_1). Moreover, by constant m_{eff} , it can be seen that with larger N_F , the maximum influence of the second order intermodulation products (at f_1) becomes practically independent on number of tones (as N_F grows, plots of the same color come closer together at the low fundamental frequencies). The reason for such behavior lies in the fact that, even though with a larger number of fundamental tones at the input, the number of distortion contributions at the output grows, their impacts decrease, because of the lower m_n . The increase of the number of distortions depending on N_F is given in the Fig. 4.7. For $N_F \geq 4$, the maximum number of second and third distortions that can fall on one tone is given by $N_F - 1$ and $N_F(N_F/2 - 1)$, respectively. On the other hand, the amplitude of one tone decreases as $A_n \sim 1/\sqrt{N_F}$ and, the impact of each second and third order distortion product, in terms of electrical power, reduces with N_F^2 and N_F^3 , respectively. This can be concluded by recalling (4.12), even though here only the expressions for harmonic distortions are given. For the impact of the second order intermodulation on the first fundamental tone, we can hence write

$$\begin{aligned} \frac{P_{\text{el,out}}(\text{IMP2})}{P_{\text{el,out}}(F)} &= 10 \log \frac{1}{4} \frac{S_2^2 (N_F - 1) \frac{A_{\text{eff}}^4}{N_F^2}}{\frac{A_{\text{eff}}^2}{N_F}} \\ &\rightarrow 20 \log \frac{S_2 A_{\text{eff}}}{2S} \quad \text{for } N_F \rightarrow \infty, \end{aligned} \quad (4.14)$$

where $A_{\text{eff}} = m_{\text{eff}} \langle I \rangle$ and which agrees with the results in Fig. 4.6.

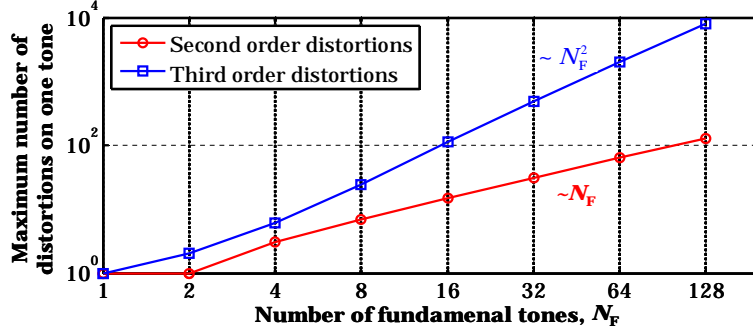


Figure 4.7: Maximum count of the second and the third order distortion products that can fall on the same frequency (separately).

Figure 4.8 a presents the minimum C/I ratio which occurs on some fundamental tone (i.e., lowest fundamental tone since it is the worst case) for different numbers of tones at the input. In a different presentation, this figure shows the reciprocal values of the points corresponding to f_1 in Fig. 4.6 (however, for two additional m_{eff} parameters and one additional N_F). In each plot, parameter m_{eff} is constant. A decrease in $(C/I)_{\text{min}}$ is observable for relatively small N_F values, however as $(N_F - 1)N_F$ approaches N_F^2 (see (4.14)), $(C/I)_{\text{min}}$ remains practically constant. From Fig. 4.8 a, a difference of some 4 dB between 2 and 128 tones can be observed, for all regarded cases of m_{eff} .

As in Fig. 4.6, it is also clear that the influence of distortions is stronger with a higher m_{eff} , and that $(C/I)_{\text{min}}$ decreases. However, even for $m_{\text{eff}} = 0.6$, $(C/I)_{\text{min}} > 21$ dB, regardless of the number of tones. Figure 4.8 b provides the maximum effective modulation depths if a minimum C/I of 20, 25 and 30 dB is desired. It can be seen that $m_{\text{eff}}^{\text{max}}$ does not depend much on the number of tones (for the same reason for which the C/I remains constant in Fig. 4.8 a), but decreases significantly (from ≈ 0.66 to ≈ 0.23) when the $(C/I)_{\text{min}}$ requirement rises from 20 to 30 dB.

In this thesis $(C/I)_{\text{min}}$ of 20 dB is regarded as sufficient. With such C/I requirement, as long as $m_{\text{eff}} \leq 0.66$ (with $N_F = 64$ for example), the maximum (second order) distortion is 100 times lower than the power of the lowest fundamental.

Consequences on system design

With the modulating signal power set to comply with an acceptable degree of linearity, the overall system performance could be enhanced by considering an array of laser diodes, all modulated with the same signal. Since one investigated VCSEL radiates an average optical power of 5 mW (at $\langle I \rangle = 40$ mA, see Fig. 4.4),

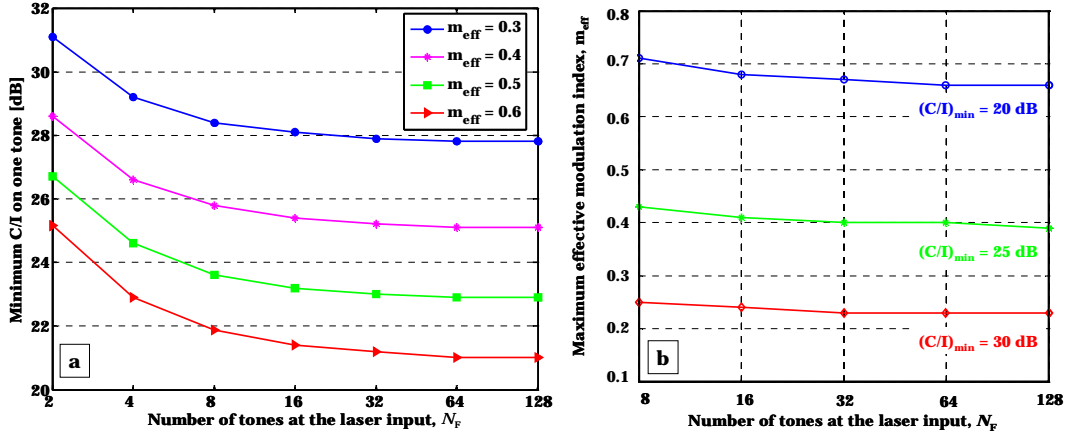


Figure 4.8: a) Minimum C/I appearing on the fundamental tone frequencies when the number of tones is varied, for different cases of signal modulation index. b) Maximum effective modulation index for different cases of minimum C/I ratio required.

the maximum number of the components within the array is determined by the safety criterium. From the discussion concerning laser safety, if an imaging diffusor renders a source with Lambertian radiation characteristic (and apparent diameter of 10 mm when set in the appropriate measurement geometry), the safety limit at the considered near-IR wavelength is about 400 mW, leading to some

$$K_{\text{LD}} = \left\lfloor \frac{P_O}{\langle P \rangle} \right\rfloor = \left\lfloor \frac{400}{5} \right\rfloor = 80 \quad (4.15)$$

devices in the array, which is a realistic figure. The operator $\lfloor \cdot \rfloor$ denotes the first integer smaller than the argument.

Additionally, the increase of the mean optical power due to the second order nonlinearities (see Fig. 4.2) is evaluated. Irrespective of N_F , when the nonlinearities are regarded, total DC optical power would be

$$\langle P \rangle = 5 \text{ mW} + (1/4)S_2 (m_{\text{eff}} \langle I \rangle)^2. \quad (4.16)$$

The influence grows with the modulation depth regarded. In the case of $m_{\text{eff}} = 0.6$, (4.16) renders $\langle P \rangle \approx 5.3$ mW. This could lead to a correction of the allowed number of components in the array (from 80 to 75). Nevertheless, since the 400 mW limit may be lifted if the immediate contact with the access point is made impossible, this issue should not be seen as very stringent.

From the conducted investigations, it can be concluded that, as far as the linearity is concerned, having a large number of tones presents a minor issue, as long as the modulating signal power does not exceed a maximum value defined by an acceptable C/I ratio. With this condition fulfilled, the system transmission rate is

primarily limited by the safety constraint and the errors which occur due to the noise or clipping of high peaks, if such is permitted at the optical source. System performance with clipping is investigated in Chapter 6.2.

In the case of the adaptive system considered in this thesis, N_F varies dynamically with the channel state. The limit on the modulating signal power is given by the required $(C/I)_{\min}$ in the worst case scenario, i.e., when all subcarriers are active. Nevertheless, as previously stated, there is no significant difference in m_{eff}^{\max} when the number of the tones is varied (it is primarily limited by the required C/I ratio). Therefore, the choice for the number of subcarriers in the system (and the FFT size) should result from the trade-off between the needed robustness against multipath dispersion and spectral efficiency (i.e., the data part of an OFDM symbol needs to be sufficiently longer than cyclic the prefix), as discussed in Chapter 6.

4.2 Infrared receiver

As schematically illustrated in Fig. 4.9, a typical optical wireless receiver consists of an optical front-end (optical concentrator and optical filter) and a photodetector (photodiode with preamplifier). An optical concentrator collects and focuses radiation, while an optical filter rejects the out-of-band ambient illumination. A photodetector converts radiation (optical power) into photocurrent (O/E conversion), before further recovering of the signal in the electrical domain.

In this subsection, attention is given to each of these components. Their principal functionalities are explained and the influence on system design is discussed, with a goal to arrive at a realistic receiver design with parameters which allow broadband transmission. However, since the focus of this thesis is to evaluate the potentials of a new signal processing technique applied to the OW channel, the least complex single element receiver design will be chosen for further investigations.

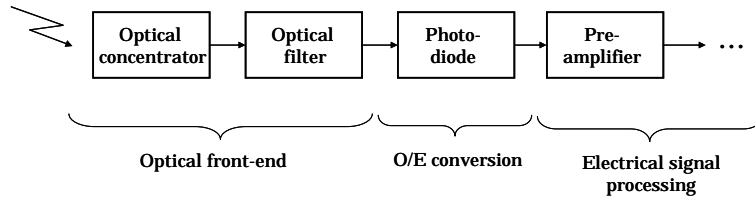


Figure 4.9: Chain of the relevant elements comprising a typical OW receiver up to the first stage of electrical signal processing.

4.2.1 Optical concentrator and optical filter

In this subsection, an overview regarding the relevant optical concentrator and optical filter design parameters is made, based primarily on the results of [41] and [72]. The discussions intend to provide a set of values favorable for the application considered.

The performance of the optical wireless link (i.e., the SNR) improves with the amount of the received optical power which is directly proportional to the effective detection area. As already indicated by (2.10) in Chapter 2.3.1, the effective surface of a bare planar photodetector ($\text{FOV}_{\text{det}} = \pi$) with an active area $A_{\text{PD}} = r_{\text{PD}}^2 \pi$ is given by

$$A_{\text{eff}}(\psi) = \begin{cases} A_{\text{PD}} \cos \psi, & 0 \leq \psi \leq \pi/2, \\ 0, & \text{otherwise,} \end{cases} \quad (4.17)$$

where ψ is the angle of incidence of the incoming light ray with respect to the detector orientation axis (a unit vector perpendicular to the detector surface, see Fig. 2.4). Hereby, in (4.17) the reflection losses are ignored. Enlarging the active area of the photodiode is quite expensive and, more importantly, leads to an increase of its capacitance, which in turn decreases the receiver bandwidth and increases the receiver noise. Dimensioning of the photodetector is addressed in the next subsection. Therefore, it is common to deploy a sort of an optical antenna (concentrator), which enlarges the effective detection area (and thus the optical power density) to

$$A_{\text{eff}}(\psi) = \begin{cases} g(\psi)T(\psi)A_{\text{PD}} \cos \psi, & 0 \leq \psi \leq \Psi_C/2, \\ 0, & \text{otherwise,} \end{cases} \quad (4.18)$$

where $g(\psi)$ and $\Psi_C \leq \pi$ are the concentrator gain and field-of-view, respectively. Factor $T(\psi)$ denotes the transmission coefficient of the optical filter. $T(\psi)$ includes all losses due to reflections and it may be calculated as an average of filter transmission at different wavelengths and/or different angles of incidence. The optical filter is discussed later on in the text and, for now, $T(\psi) = 1$ is assumed.

The concentrator can be of non-imaging or imaging type, where the former is usual for the short-range links considered in this work. A theoretical maximum of a non-imaging concentrator gain is determined by the invariant *etendue*¹¹ [73]. According to this fundamental principle, the theoretical maximum concentrator gain of a 3D non-imaging optical concentrator (concentrating the radiation from the air to the photodetector) is

$$g_{\text{max}} = \frac{n^2}{\sin^2(\Psi_C/2)}, \quad (4.19)$$

¹¹Etendue (French for extent space) presents a constant in the optical system, which is used to specify its geometric capability to transmit radiation, i.e., it determines the power flow (throughput). It is also referred to as *extent*, *acceptance*, *Lagrange invariant*, etc.

where Ψ_C denotes the concentrator field of view and n is the refraction index of the concentrator material (glass). A more detailed derivation of (4.19) is presented in Appendix D. Clearly, there is a trade-off between the concentrator gain and its FOV. Whether the gain (4.19) can be achieved in practice depends on the concentrator design and dimension. 3D concentrators which have spherical symmetry in their construction (referred to as *spherical truncated concentrators*) as well as some with cylindrical symmetry (*compound parabolic concentrators*) are able to reach it. Both types of concentrators are widely used in optical wireless links.

Since a non-directed link is considered in this work, a receiver with a relatively large FOV is desirable. A simple and inexpensive solution is to use a so-called *hemispherical concentrator*, shown in Fig. 4.10 (a and b). Hemispherical concentrator is a spherical truncated concentrator with a truncation angle $\theta_t = \pi/2$. If correctly dimensioned (i.e., if the concentrator radius R_C exceeds the detector radius r_{PD} by a factor n^2 , where n denotes its refraction index), an idealized hemispherical concentrator (i.e., with neglected reflection and filter losses) achieves a wide FOV ($\Psi_C \approx \pi$) with practically constant gain of n^2 [41]. Compared to the hemispherical concentrator, a compound parabolic concentrator can provide significantly higher gains, but due to the comparatively smaller FOVs, it is more attractive for directed links [51]. A wide FOV receiver can be realized by an array of compound parabolic concentrators, oriented in different directions [25]. Such approach is much more complex, and requires deployment of a photodetector array, which remains outside the scope of this work.

Almost all applications require deployment of optical filters to reject the out-of-band ambient light radiation. Relatively simple high-pass filters made of colored glass or plastic are usual in commercial IR systems so far. Such filters pass the radiation beyond a certain wavelength, and in combination with the spectral response of a Si-photodetector, quite a wide pass-band (hundreds of nm) is realized. Alternatively, the use of direct gap semiconductors in combination with a Si-detector was proposed in [74], where a pass-band of 100 nm was reported as conceivable. A newer, more complex type is the thin-film band-pass optical filter which can have far narrower bandwidths (< 1 nm) and thus achieve a superior rejection of ambient light. Because of the very narrow passband, laser diode is the only suitable optical source for combination with thin film filters. An important feature of these pass-band filters is that their transmission characteristic depends on the angle of incidence, which is not the case for the high-pass filters. For relatively small angles of incidence ($\theta \approx 30^\circ$), only the center wavelength shifts towards shorter wavelengths (about 2%), but the spectral shape of the filter characteristic remains unchanged, whereas for $\theta \approx 60^\circ$ the spectral characteristic also transforms considerably. Therefore, for wide FOV applications, a passband filter needs to be carefully designed, i.e., enough bandwidth should be allowed to take into account the mentioned effects. Moreover, complexity in the fabrication of these filters is much higher than

in the case of high-pass filters.

Assuming a hemispherical concentrator, the optical filter could be placed either between the concentrator and the detector (planar filter, Fig. 4.10 a) or above the concentrator (hemispherical filter, Fig. 4.10 b). A planar thin film filter is, however, not convenient in combination with a wide FOV receiver. The range of incident angles on the planar filter is the same as on the detector. Thus, the angular filter bandwidth (i.e., filter FOV, θ_C) needs to cover all possible incoming angles of incidence, which in case from Fig. 4.10 a means $\theta_C = \pi$. In [72], an approximative relationship between the filter FOV and its bandwidth $\Delta\lambda$ was given

$$\sin \frac{\theta_C}{2} \approx \frac{n_S}{n_1} \sqrt{\frac{\Delta\lambda}{\lambda}}, \quad (4.20)$$

where n_1 denotes the refractive index of the input layer (concentrator glass for the planar filter, see Fig. 4.10 a) and $n_S = 2.293$ is the empirically obtained effective refractive index of the spacer layer of a typical filter structure proposed in [41].¹² According to (4.20), it can be concluded that large θ_C leads to the requirement of a large filter bandwidth. For example, for $\theta_C = \pi$, $n_1 = 1.8$, and $\lambda = 850$ nm, the filter would need a prohibitively large bandwidth of 524 nm.

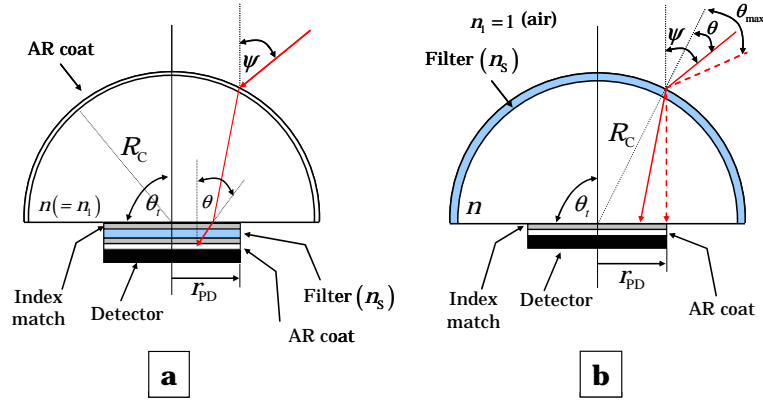


Figure 4.10: Structure of an OW receiver front-end deploying a hemispherical concentrator in combination with a) planar and b) hemispherical optical filter. AR stands for anti-reflection coating.

The situation is different for the hemispherical filter (Fig. 4.10 b). No matter how large the angle ψ of the incoming light ray with respect to the detector surface

¹²Relationship (4.20) was obtained under the assumption of a Butterworth filter transmission characteristic (comparisons with actual characteristics can be found in [72]). The approximation is accurate when the filter center wavelength coincides with the signal wavelength by normal incidence (i.e., when no center wavelength shift occurs by normal incidence). Quantity $n_S = 2.293$ was empirically obtained in [72] to excellently fit an analytical expression for the center wavelength shift.

is, the angle of incidence to the hemispherical filter θ (see Fig. 4.10 b) is always confined to the interval $[0, \theta_{\max}]$, determined by

$$\theta_{\max} = \arcsin\left(\frac{nr_{\text{PD}}}{R_C}\right). \quad (4.21)$$

This means that by choosing an R_C large enough, θ_{\max} can be made sufficiently small so that the angle dependence of the spherical filter is acceptable. In fact, this means that it is possible to arbitrarily chose the filter bandwidth (which dictates a certain θ_C by (4.20)) and then, with proper dimensioning of concentrator radius, to force $\theta_{\max} \leq \theta_C/2$. The needed R_C for a given $\Delta\lambda$ can be expressed as

$$R_C \geq r_{\text{PD}} n \frac{n_1}{n_S} \sqrt{\frac{\lambda}{\Delta\lambda}}. \quad (4.22)$$

For a hemispherical filter $n_1 = 1$ (input layer is air, see Fig. 4.10 b). For example, if a hemispherical filter with $\Delta\lambda = 15$ nm is desired in combination with a detector whose active area is $A_{\text{PD}} = r_{\text{PD}}^2 \pi = 1$ cm², the influence of the filter angle of incidence could be made negligible by using a glass concentrator ($n \approx 1.8$) with a radius $R_C \approx 3.3$ cm. Nevertheless, for portable terminals, size and weight limitations can restrict the choice of the concentrator radius.

Another effect, which can dramatically reduce the gain of the idealized concentrator, especially for non-normal incident angles, are the reflections occurring at the concentrator input, at its output, and at the input of the detector. Their influence can however be significantly reduced by careful placement of antireflection (AR) coatings and index matching compounds, as illustrated in Fig. 4.10. As already mentioned, the losses due to reflections as well as filter transmission are included in the factor $T(\psi)$.

As an illustration of previous discussions, a result of the investigations reported by Barry in [41] is given in Fig. 4.11. This figure shows a comparison of the hemispherical concentrator optical gain performance, depending on the angle of incidence (with respect to the detector orientation), for the case when all reflection and filter losses are neglected (ideal, $T(\psi) = 1$), for the cases when reflection losses are taken into account, with AR coatings and without (i.e., when space between concentrator and detector is filled with air), and for the cases when a hemispherical and planar optical filter is used (as shown in Fig. 4.10). In the shown example of [41], a concentrator with radius $R_C = 2$ cm and refraction index $n = 1.8$, and a circle-shaped detector with active area $A_{\text{PD}} = 1$ cm² were assumed. Filters were modeled with 3rd order Butterworth characteristic with peak transmission¹³

¹³The peak transmission factor is determined by technology, and should be chosen as large as possible. Usually, it takes values between 0.4 and 0.9.

$T_0 = 0.7$. Same filter bandwidths of $\Delta\lambda = 15$ nm were taken and orientations¹⁴ of 45° for the planar and of 15° for the hemispherical filter were assumed, by a nominal signal wavelength $\lambda = 810$ nm.

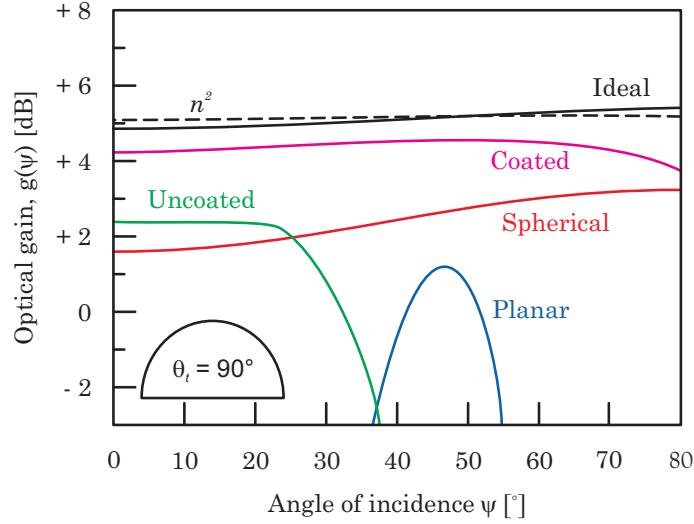


Figure 4.11: Optical gain of a hemispherical concentrator depending on the incoming light ray angle of incidence [41]. “Ideal” plot considers no losses; “Coated” and “Uncoated” consider the structure with and without AR coating, respectively; “Spherical” and “Planar” consider structures with hemispherical and planar optical filter, respectively.

For the idealized concentrator, the maximum gain (n^2) is achievable practically over the full FOV, and varies only slightly (within 0.3 dB from the ideal gain) with the angle. Note, however, that the hemispherical concentrator is in that way special and this is not the case for other types of spherical truncated concentrators. By comparing the curves labeled as uncoated and coated, it is clear that reflections evoke a narrow FOV, unacceptable for non-directed applications, but that with deployment of AR coatings and index matching compounds, the gain stays within 1.5 dB from the ideal, even for extremely large angles of incidence. Furthermore, when two types of filters were compared, it can be observed in the figure that the gain is practically omnidirectional with the spherical filter, whereas the system with the planar filter exhibits any gain only for a narrow FOV around 45° . Clearly, even though both types of filter equally effectively reject the ambient light, only the hemispherical filter is suitable for wide FOV applications.

¹⁴The filter orientation is defined as the filter angle of incidence for which there is no shift of the filter center wavelength.

4.2.2 Photodetector and preamplifier

The photodetector represents the main component of an OW receiver, which performs conversion of light into electrical current. As in fibre-optic applications, it should exhibit high sensitivity, fast response, low noise, and low cost. However, different than in fibre-optic applications, where the size of the detector needs to be compatible to the small fibre core, large-area devices are required here to collect enough signal energy which is in a non-directed wireless system spread over a wide region. Low-cost, large-area Si-PDs, whose peak responsivity (@ ≈ 900 nm) coincides with the near IR transmission window are commercially available. However, broadband communication is of interest in this thesis, and the desired large electrical receiver bandwidth is not easily achievable with a large-area detector, due to its higher capacitance. In Chapter 2, it has been observed that the purely diffuse channel has a bandwidth of about 10 MHz, but that the LOS channel can have a much higher bandwidth. The aim of the following discussions is therefore to characterize a photodetector which would be able to support broadband transmission, i.e., to detect 100 MHz wide baseband signals.

Both PIN and avalanche photodiodes (APD) are available in Si-technology, and both have been used in OW systems with single element detector. The comparison of their main qualities is the same as in fibre-optics - APDs have superior sensitivity, while PINs are much simpler and less expensive. If an OW link with PIN is already limited by shot noise due to background radiation, deployment of an APD leads to a decrease of link SNR. On the other hand, in the case where preamplifier noise is dominant, using an APD may significantly enhance the sensitivity and extend the link reach. Based on the mentioned requirements, in the following text, investigations and discussions are conducted towards the choice for PD type and dimension, as well as type of preamplifier, suitable for deployment in the considered system.

The issue of the transmission wavelength choice was already discussed from the optical source point of view in the Subsection 4.1.1, where the laser safety was regarded. There, the short wavelength extended sources were found as preferable. As far as detectors are considered, photodiodes suitable for the IR transmission in 1300 nm and 1550 nm windows (conventionally used in fibre-optical telecommunications) are made of expensive materials (InGaAsP and InGaAs [75]) and more importantly, they are characterized by a quite large capacitance per unit area (about 30 pF/mm² compared to Si which has 1 pF/mm² [76]). The main reason for their high capacitance per unit area lies in the small width of the depletion region w (which is inversely proportional to the capacitance, as shall be seen later). With materials like InGaAsP and InGaAs, a depletion region of only a few μm is sufficient to achieve an excellent quantum efficiency. At the same time, such thin depletion region minimizes the carrier transit time, which is, in fiber-optic systems,

of greater importance than the capacitance per unit area (since detectors have a small surface anyhow). As a consequence, using this λ -range is not the best option for large area devices, especially for broadband applications. On the other side, the Si-based photodiodes need a depletion region of several tens of μm to realize reasonable quantum efficiencies, resulting in much lower capacitance per unit area.

Photodetector bandwidth

The *bandwidth* of a photodetector is determined by the speed with which it responds to the variations of the incident optical power. There are two effects which limit the bandwidth of a PIN detector - the RC constant because of the electric parasitics and the so-called transit time τ_{tr} , which presents the period between the carriers' generation (by photon absorption) and their collection. For the upcoming discussions, the RC -circuit bandwidth and transit-time bandwidth are individually defined as [77]

$$f_{\text{RC}} = (2\pi R_{\text{in}} C_{\text{TOT}})^{-1} \quad (4.23)$$

and

$$f_{\text{tr}} = \frac{0.44}{\tau_{\text{tr}}}, \quad (4.24)$$

respectively. In the case of an APD detector, apart from the transit time related to the depletion layer, there is an additional effective transit time τ_{M} within the multiplication layer. This transit time influences the detector bandwidth as

$$f_{\text{M}} = (2\pi\tau_{\text{M}}M_0)^{-1}, \quad (4.25)$$

where M_0 denotes the multiplication factor at low frequencies. The bandwidth of an APD photodetector becomes

$$f_{\text{g}}^{-1} = f_{\text{RC}}^{-1} + f_{\text{tr}}^{-1} + f_{\text{M}}^{-1}, \quad (4.26)$$

and for a PIN photodiode, the last summand is omitted.

First, the *RC-circuit bandwidth* is discussed. Preamplifiers can be designed as low-impedance (typically $50\ \Omega$), high impedance and transimpedance (TIA) amplifier. Each type is associated with features which are differently appreciated, depending on the application.

The simplified model scheme for a detector with low- and high-impedance preamplifier is shown in Fig. 4.12 a [75], and the equivalent small signal circuit in Fig. 4.12 b [77]. The detector is usually AC-coupled to avoid saturation of the amplifier by the background light.

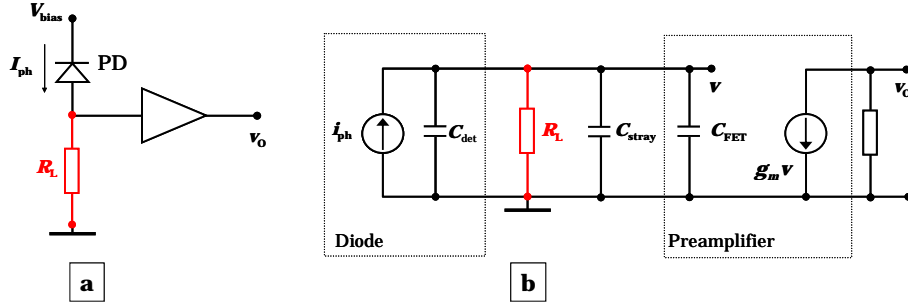


Figure 4.12: a) Simplified scheme of a detector with low- or high-impedance preamplifier [75], b) Equivalent small circuit scheme without noise sources [77].

The low- or high- impedance preamplifier design is the most straightforward one, with the input of the preamplifier terminated by an external load resistor, R_L . Thus, the current from the reverse-biased photodiode creates a voltage across R_L , which is fed into the preamplifier input transistor. The RC -circuit bandwidth is determined by the total resistance and total capacitance at the input of the preamplifier transistor. The diode can be equivalently represented by a current source in parallel to the depletion capacitance, C_{det} (shunt capacitance)¹⁵. Parasitic effects of the interconnection between the diode and preamplifier are accounted for by the stray capacitance, C_{stray} . The gate capacitance of the field-effect transistor (FET) used for amplification is included as C_{FET} . Realization of the preamplifier circuit with FET transistors is assumed in all discussions due to their superior noise performance compared to bipolar transistors [75]. Since the input admittance of a FET is capacitive [78], we can write $R_{in} = R_L$. Therefore,

$$f_{RC} = \frac{1}{2\pi R_{in} C_{TOT}} = \frac{1}{2\pi R_L (C_{det} + C_{stray} + C_{FET})}. \quad (4.27)$$

In Fig. 4.12 b, the receiver circuit is shown without noise sources. The noise analysis is conducted later on, when the receiver sensitivity is discussed.

For low-impedance receivers, $R_L = 50 \, \Omega$ is typically chosen, allowing for wide receiver bandwidth (the same as signal bandwidth). On the other hand, as shall be seen from noise analysis, a small R_L boosts the thermal noise, reducing so the receiver sensitivity.

An opposite situation is encountered in case of the high-impedance preamplifier. Large values of R_L (in the order of magnitude 10 k Ω) are useful for receiver sensitivity, but the resulting bandwidth could be prohibitively small for high-speed transmission.¹⁶ The use of an equalizer after the preamplifier is not very attrac-

¹⁵Shunt resistance and conductance are both small, with negligible effect on the performance.

¹⁶Even in fibre-optic communications, where the input capacitance is much smaller, about 1 pF, a 10 k Ω input resistance would lead to an $f_{RC} \approx 16$ MHz.

tive, because it enhances the noise at higher frequencies. Moreover, the pole of the frequency response can vary with age and temperature. The high-impedance receiver is, therefore, not further regarded in this work.

The TIA preamplifier design is presented in Fig. 4.13, where again a simplified diagram is shown (Fig. 4.13 a) [75], together with the equivalent small circuit scheme (Fig. 4.13 b) [77].

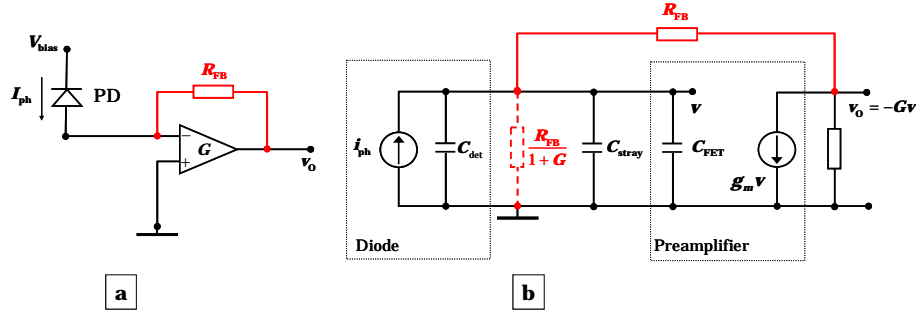


Figure 4.13: a) Simplified scheme of a detector with transimpedance preamplifier [75], b) Equivalent small circuit scheme without noise sources [77].

This receiver uses a large feed-back resistor R_{FB} connected around an inverting amplifier. If the voltage gain of the amplifier is denoted by G , the negative feed-back ensures that the input resistance is decreased as

$$R_{in} = \frac{R_{FB}}{1 + G}. \quad (4.28)$$

In this way, receivers with TIA preamplifier can achieve high sensitivity together with large bandwidth. The expression for C_{TOT} remain the same as in (4.27).

While the preamplifier design determines the receiver input resistance, its input capacitance is mainly influenced by the depletion capacitance of the large-area photodiode. Typically, C_{stray} has a pF order of magnitude, and the preamplifier capacitance C_{FET} is negligible due to the small length and width of the transistor gate. In [41], $C_{FET} = 0.1(C_{det} + C_{stray})$ was recommended as a good trade-off between the cost and capacitance increase, and is here also assumed. The depletion capacitance of a planar detector is determined by the material, the depletion region width (i.e., the width of the intrinsic layer w) and the detector area as

$$C_{det} = \varepsilon_0 \varepsilon_r \frac{A_{PD}}{w}, \quad (4.29)$$

where $\varepsilon_0 = 8.85 \cdot 10^{-12} \text{ Fm}^{-1}$, is the vacuum permittivity and ε_r is the relative permittivity of the particular medium, which is for silicon $\varepsilon_r(\text{Si}) = 11.7$ [78]. Equation (4.29) reveals the relation between capacitance per unit area and w , mentioned in

the discussion of photodiodes operating at longer wavelengths. Clearly, with the increase of the detector area, its input capacitance grows, reducing in turn the detector bandwidth. According to (4.29), the input capacitance could possibly be reduced with a thicker depletion region.

However, the following discussion of the *transit-time bandwidth* clarifies that too thick w is not recommendable. Even though a thick i -zone decreases the input capacitance, it also increases the transit time, since the carriers need more time to drift across the depletion region. The transit time is defined as [64]

$$\tau_{\text{tr}} = \frac{w}{v_d}, \quad (4.30)$$

where v_d is the saturation drift velocity of the minority carrier in the illuminated layer. Since holes travel slower than electrons, the more critical case for bandwidth would be to illuminate the n contact, so then $v_d \approx 10^5 \text{ m/s}$. From this aspect, large w has detrimental influence on the detector bandwidth as [77]

$$f_{\text{tr}} = \frac{0.44v_d}{w}. \quad (4.31)$$

As already mentioned, the width of the i -region, however, does not influence only the receiver bandwidth, but its efficiency as well. Namely, while carrier pairs generated in the depletion region rapidly drift across it under the influence of high voltage applied, they travel much slower via diffusion in the doped layers (p or n). For a decrease of the detector response time it would therefore be of benefit to have thin p and n layers, and a depletion layer wide enough so that all incident photons are absorbed while traversing it. In the case of Si semiconductors, typically $w \approx 20\text{-}50 \mu\text{m}$ ensures a reasonable quantum efficiency (i.e., responsivity). For such range of w values, (4.31) leads to bandwidths of $0.88 - 2.2 \text{ GHz}$.

The avalanche photodiode has an additional layer - the so called *multiplication* layer - in which secondary carrier pairs are generated through impact ionization. The intrinsic layer still acts as the depletion region in which most of the incident photons are absorbed and primary carrier pairs are generated. Electrons generated in this region cross the multiplication layer and generate secondary pairs responsible for the current gain. Thus, while the responsivity of a PIN can not be larger than 1, the responsivity of an APD is increased because of the current gain by the multiplication factor M (average). That increases the electrical power of the detected signal by M^2 . Multiplication factor is dependent on the impact-ionization coefficients of electrons (α_e) and holes (α_h), as well as on the thickness of the multiplication region [64]. Because of the additional generation and collection of the secondary carrier pairs, a longer time is required to move a unit distance within the multiplication layer than outside of it. The transit time within the multiplication layer τ_M influences the detector bandwidth as introduced by (4.25). The product

of bandwidth and amplification is constant, i.e., there is a trade-off between speed and sensitivity. With a typical value of $\tau_M = 0.8$ ps for Si, $M_0 f_M \approx 200$ GHz [79] is obtained, with best values for this product in the range of 300 GHz [77]. This multiplication time might therefore be a problem when the APD is used at a gain of several hundred. Generally, $M(f) \approx M_0$ can be assumed. Multiplication factor is hence from now on denoted simply by M .

Noise analysis and receiver sensitivity discussion

Two fundamental *noise mechanisms* in optical receivers are *shot noise* and *thermal noise*. These lead to fluctuations of the photocurrent even when the incident optical signal has a constant power, hereby affecting the receiver performance.

Definition of shot noise has already been introduced in Chapter 2.4, but is repeated for the sake of completeness and consistency. Shot noise arises because the electrons forming the photocurrent are generated at random instances of time. The current fluctuation related to the shot noise is a stationary random process, which is in practice (in case of high intensity) approximated by a Gaussian distribution with the variance [64]

$$\sigma_S^2 = 2qI_{pc}B, \quad (4.32)$$

for the PIN-based receiver, where $q = 1.6 \cdot 10^{-19}$ [As] is the electron charge, $I_{pc} = \eta P_{Rx}$ is the average photocurrent generated by the signal of constant optical power P_{Rx} detected by the receiver of responsivity η . B is the effective noise bandwidth of the receiver (corresponding to the intrinsic photodetector bandwidth, with a value depending on the receiver design).

Thermal noise is a manifestation of the fact that, at a finite temperature, electrons move randomly in a conductor. This generates photocurrent fluctuations in resistors even without application of voltage. These fluctuations can be modeled as a stationary Gaussian random process, which is practically independent on frequency (nearly white noise), with the variance given by

$$\sigma_T^2 = \frac{4k_B T}{R} B, \quad (4.33)$$

where $k_B = 1.38 \cdot 10^{-23}$ Ws/K is the Boltzmann constant, T is the absolute temperature and R denotes the load resistor in the front-end ($R = R_L$ for the low impedance and $R = R_{FB}$ for the transimpedance preamplifier).

The two random processes are independent and practically Gaussian, so the total noise variance can be calculated as $\sigma^2 = \sigma_S^2 + \sigma_T^2$. Note that in the general case the thermal noise does not depend on the average photocurrent I_{pc} , whereas the

shot noise does. This is also the reason why APD receivers suffer from shot noise enhancement. Since the secondary carrier pairs are generated in the multiplication region at random times, it causes an additional contribution of the shot noise. The shot noise variance (4.32) accordingly modified for an APD receiver is

$$\sigma_S^2 = 2qM^2F(M)I_{pc}B, \quad (4.34)$$

where $F(M) = k_A M + (1 - k_A)(2 - 1/M)$ is the *excess noise factor* of the APD. The quantity $k_A = \alpha_h/\alpha_n$ is the ionization coefficient ratio in a silicon photodiode. A small k_A is clearly advantageous and silicon typically has $k_A \ll 1$. Since all corresponding expressions for the PIN photodiode can be obtained by setting $M = 1$ and $F(M) = 1$, the more general forms are used from now on. Even though the power of the detected signal enhances M^2 times, because of the excess noise, the SNR improvement by an APD is considerably reduced from M^2 . In shot-noise dominant environments, such as the OW link with strong ambient light, this leads to a decrease of SNR with respect to the PIN photodiode.

Figure 4.14 presents the equivalent circuit with *noise sources* for the optical receivers considered in Fig. 4.12 and 4.13.

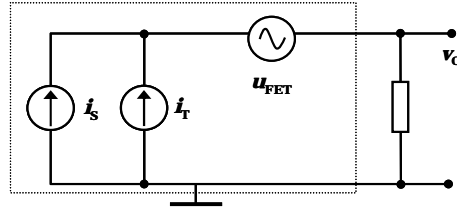


Figure 4.14: Scheme of the equivalent circuit with noise sources in detector. Signal current source is left out.

The AC current i_s denotes fluctuations related to the shot noise. Shot noise is generated by the incident optical power stemming from both the signal (related to $\sigma_{S,\text{sig}}^2 \sim I_{\text{sig}} \sim P_{\text{sig}}$) and the background light (related to $\sigma_{S,\text{BG}}^2 \sim I_{\text{BG}} \sim P_{\text{BG}}$). The dark and the gate leakage current also contribute to receiver shot noise but are usually neglected in the analysis. In typical OW environments, the in-band background light power at the receiver is much stronger than the signal power ($P_{\text{BG}} \gg P_{\text{sig}}$) [41], so the shot noise component of an OW receiver becomes

$$\langle i_s^2 \rangle = \sigma_{S,\text{BG}}^2 = 2qM^2F(M)I_{\text{BG}}B, \quad (4.35)$$

where $I_{\text{BG}} = \eta P_{\text{BG}} = \eta p_{\text{BG}} \Delta \lambda A_{\text{eff}}$. Hereby, a background irradiance $p_{\text{BG}} = 5.8 \mu\text{W}/\text{nm}/\text{cm}^2$ corresponding to the bright sky without direct sun exposure [60] is always assumed (as worst case).

The current i_T is related to the thermal noise contribution on the load or feedback resistor

$$\langle i_T^2 \rangle = \sigma_T^2, \quad (4.36)$$

in the case of the low-impedance and trans-impedance preamplifier circuits, defined in (4.33). Both $\langle i_S^2 \rangle$ and $\langle i_T^2 \rangle$ depend linearly on the detector bandwidth.

Noise of the FET-based preamplifier is described as the voltage source u_{FET} with

$$\langle u_{\text{FET}}^2 \rangle = \Gamma \frac{4k_B T}{g_m} B, \quad (4.37)$$

where Γ is the FET channel noise factor and g_m is its transconductance [77]. The spectral density of the equivalent current source variance is

$$\frac{d \langle i_{\text{FET}}^2 \rangle}{df} = \frac{d \langle u_{\text{FET}}^2 \rangle}{df} \left(\frac{1}{R^2} + (2\pi f C_{\text{TOT}})^2 \right), \quad (4.38)$$

so that the FET variance over the effective bandwidth B results in

$$\sigma_{\text{FET}}^2 = \Gamma \frac{4k_B T}{g_m} \frac{1}{R^2} B + \frac{1}{3} \Gamma \frac{4k_B T}{g_m} (2\pi C_{\text{TOT}})^2 B^3. \quad (4.39)$$

This leads to the total receiver noise

$$\sigma^2 = \sigma_{S,BG}^2 + \sigma_T^2 + \sigma_{\text{FET}}^2 \quad (4.40)$$

in $[A^2]$, and to the signal-to-noise-ratio

$$\begin{aligned} \text{SNR} &= \frac{\eta^2 M^2 P_{\text{sig}}^2}{\sigma_{S,BG}^2 + \sigma_T^2 + \sigma_{\text{FET}}^2} \\ &= \frac{\eta^2 M^2 P_{\text{sig}}^2}{2qM^2 F(M) I_{BG} B + \frac{4k_B T}{R} B + \Gamma \frac{4k_B T}{g_m} \frac{1}{R^2} B + \frac{1}{3} \Gamma \frac{4k_B T}{g_m} (2\pi C_{\text{TOT}})^2 B^3}. \end{aligned} \quad (4.41)$$

Note that the FET noise has a term which depends on B^3 . Also, due to the large detector size, the capacitance C_{TOT} is much larger than in fibre-optic applications.

Besides bandwidth, receiver sensitivity is a very important criterium for system design. It is usually described by the minimum optical signal power at the receiver required for detection with an acceptable error rate. This can be expressed as

$$P_{\text{sig,min}} = \sqrt{\text{SNR}_{\text{min}}(\text{MF}, \text{BER}) \frac{\sigma^2}{M^2 \eta^2}}, \quad (4.42)$$

where MF is abbreviation for modulation format. If the expression for σ^2 , explicitly given in the denominator of (4.41), is looked at in detail, it can be concluded that,

for a given bandwidth, the shot noise rises linearly with the detector surface, the thermal noise remains unchanged, and the FET noise grows with the square of the detector surface. In case of the APD diode, the shot noise contribution is additionally enlarged by the factor $M^2 F(M)$, but the signal power is also increased M^2 times. The sensitivity of an APD receiver depends on the chosen M . For each detector design there is an optimal multiplication factor M_{opt} which provides the highest sensitivity (i.e., results in minimum $P_{\text{sig,min}}$).¹⁷ It can be obtained by solving

$$\frac{\partial \text{SNR}}{\partial M} = 0, \quad (4.43)$$

where $\text{SNR}(M)$ is given by (4.41). If we define a constant Q as a ratio

$$Q = \frac{\sigma_{\text{T}}^2 + \sigma_{\text{FET}}^2}{2qI_{\text{BG}}B}, \quad (4.44)$$

after derivation and some manipulations (4.43) can be translated into

$$M^3 + \left(\frac{1}{k_{\text{A}}} - 1 \right) M - \frac{2Q}{k_{\text{A}}} = 0. \quad (4.45)$$

The optimal multiplication factor is found as the real valued solution of a cube equation (4.45)

$$M_{\text{opt}} = \frac{1 - \frac{1}{k_{\text{A}}} + 3 \left(\sqrt{\frac{1}{27} \left(\frac{1}{k_{\text{A}}} - 1 \right)^3 + \frac{Q^2}{k_{\text{A}}^2} + \frac{Q}{k_{\text{A}}}} \right)^{2/3}}{3 \left(\sqrt{\frac{1}{27} \left(\frac{1}{k_{\text{A}}} - 1 \right)^3 + \frac{Q^2}{k_{\text{A}}^2} + \frac{Q}{k_{\text{A}}}} \right)^{1/3}}. \quad (4.46)$$

Figure 4.15 shows the optimal multiplication factors of an APD receiver, depending on its size (characterized by a diameter $d = 2r_{\text{PD}}$ and surface $A_{\text{PD}} = r_{\text{PD}}^2 \pi$). Calculations were made for both low impedance and TIA preamplifier with two different feed-back resistors. Table 4.2 summarizes the values of parameters used for calculation. Unless stated differently, these parameters are considered for the further calculations as well.

The values of feed-back resistors are typical for TIA preamplifiers. The values for the FET channel noise factor and transconductance are determined by the transistor choice - clearly with small Γ and large g_m , thermal noise would be reduced. The orders of magnitude for the values regarded here agree with [41], where the complete receiver optimization was performed.

¹⁷Practical APD receivers are configured so that the chosen value of M renders the minimum BER.

Optical front-end	Photodiode	Preamplifier
$p_{\text{BG}} = 5.8 \mu\text{W}/\text{nm}/\text{cm}^2$	$\eta = 0.7 \text{ A/W}$	$R_{\text{L}} = 50 \Omega$ (low-imp.)
$n = 1.8$ (concentrator)	$w = 20 \mu\text{m}$	$R_{\text{FB}} = 1 \text{ k}\Omega, 10 \text{ k}\Omega$ (TIA)
$T_0 = 0.8$ (filter)	$\epsilon_{\text{r}} = 11.7$ (Si)	$g_{\text{m}} = 0.1 \text{ S}$ (FET)
$\Delta\lambda = 15 \text{ nm}$ (filter)	$C_{\text{stray}} = 1 \text{ pF}$	$\Gamma = 2/3$ (FET)
	$B = 100 \text{ MHz}$	
	$k_{\text{A}} = 0.03$	

Table 4.2: Parameters for optimal multiplication factor calculations (results in Fig. 4.15) and for calculations of the noise contributions' powers (results in Fig. 4.16).

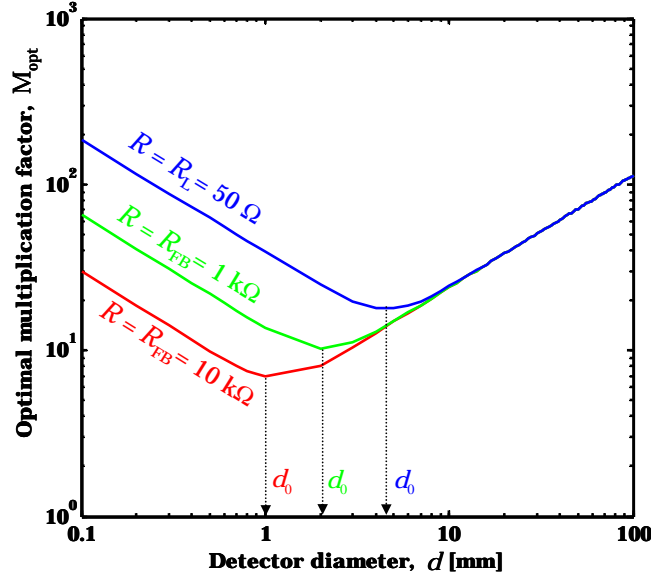


Figure 4.15: Optimal multiplication factor of the APD receiver as a function of detector diameter, d . Calculation parameters are given in Table 4.2.

By observing Fig. 4.15, it can be concluded that the value of M_{opt} clearly depends on the relationship between the factors of the constant Q , i.e., on $\sigma_{\text{T}}^2/(2qI_{\text{BG}}B)$ and $\sigma_{\text{FET}}^2/(2qI_{\text{BG}}B)$. By inspecting (4.33) and (4.39), it can be concluded that the former factor decreases with the detector diameter, while the latter one grows. The slope of the M_{opt} curve is hence determined by the more dominant of the two factors. More precisely, for small d , $\sigma_{\text{T}}^2 \gg \sigma_{\text{FET}}^2$, so $Q \sim 1/d^2$ and the optimal multiplication factor decreases as $\sim 1/d^{2/3}$ (4.46). However, for large d , when $\sigma_{\text{FET}}^2 \gg \sigma_{\text{T}}^2$, $Q \sim d^2$ and M_{opt} is increasing with diameter as $\sim d^{2/3}$. Because of the opposite slope signs of Q for small and large d , there is a minimum in the $M_{\text{opt}}(d)$ curve for $d = d_0$. Moreover, from (4.46) it can be derived that $M_{\text{opt}} \rightarrow \infty$ for both $d \rightarrow 0$ and $d \rightarrow \infty$, which substantiates the observations from Fig. 4.15.

Furthermore, it is clear that, when σ_T^2 is dominant ($d \ll d_0$), the choice of the load resistor influences M_{opt} (larger resistor leads to lower M_{opt}), whereas, when the capacitive component of σ_{FET}^2 takes over ($d \gg d_0$), the resistor size becomes irrelevant (for M_{opt} and for sensitivity). Because of such shape of the curves, it can be observed that $M_{\text{opt}}(d_0)$ diminishes with larger R . For the cases considered in Fig. 4.15, $M_{\text{opt}} > 1$, meaning that an APD would in all cases render better sensitivity than a PIN. By extension of the $\sim d^{2/3}$ parts of the curves ($d \gg d_0$), one could estimate that an $M_{\text{opt}} = 1$ would be the choice for a rather small detector ($d_0 \approx 0.15$ mm) with extremely large feed-back resistor. The sensitivity enhancement by deployment of an APD photodiode is discussed later on.

Figures 4.16 a-c show the behavior of the noise contributions (in terms of power), with the low impedance preamplifier (Fig. 4.16 a) and two TIA preamplifiers (Fig. 4.16 b and c). For the calculations, a bandwidth $B = 100$ MHz is assumed, since this can be regarded as a good value for broadband transmission (in the considered channel). The shot noise plots of the APD receiver are obtained with M_{opt} for the corresponding resistor and detector diameter (Fig. 4.15). The noise factors are presented for both APD and PIN detector.

In the considered PIN receivers, it can be seen that **1)** the shot noise component (red dotted plots) grows with d^2 and is independent on the preamplifier design; **2)** the thermal noise (green plots) is significantly reduced by deployment of the TIA preamplifier, and is independent on detector the size and **3)** the FET noise component (blue plots) is reduced for small detector sizes (where the resistive component of the impedance is dominant), but as the detector diameter grows, the capacitive component of the impedance becomes dominant, letting the noise grow with d^4 . For smaller detectors, the total noise in the PIN receiver (black dotted points) is dominated by σ_T^2 , whereas with increase of detector size, it becomes dominated by σ_{FET}^2 . Transition of the σ_{PIN}^2 slope occurs at $d = d_1$, when $\sigma_T^2 = \sigma_{\text{FET}}^2(d_1)$. Obviously, the crossing point between σ_T^2 and σ_{FET}^2 depends on the load resistance deployed.¹⁸ With larger R_{FB} , σ_T^2 decreases, as well as the resistive part of the FET impedance (so, its capacitive part becomes dominant for smaller d). As a consequence, the cross-point d_1 shifts towards smaller values. By inspection of Fig. 4.16 a-c, it can be seen that all three PIN receivers (low-impedance and both TIA) are limited by the preamplifier noise, however, with the increase of R_{FB} an area around d_1 appears (and grows) where the shot noise plays a significant role and could even become dominant for sufficiently large feedback resistors.

On the other hand, all APD receivers in Fig. 4.16 a-c are shot noise limited, since M_{opt} makes $\sigma_{\text{S,BG}}^2(\text{APD}, M_{\text{opt}}) = 2qI_{\text{BG}}M_{\text{opt}}^2F(M_{\text{opt}})B$ the prevailing noise component. It can be seen that the decrease in M_{opt} with increase of the detector

¹⁸It is assumed that the transistors and other resistors in the preamplifier circuit are fixed.

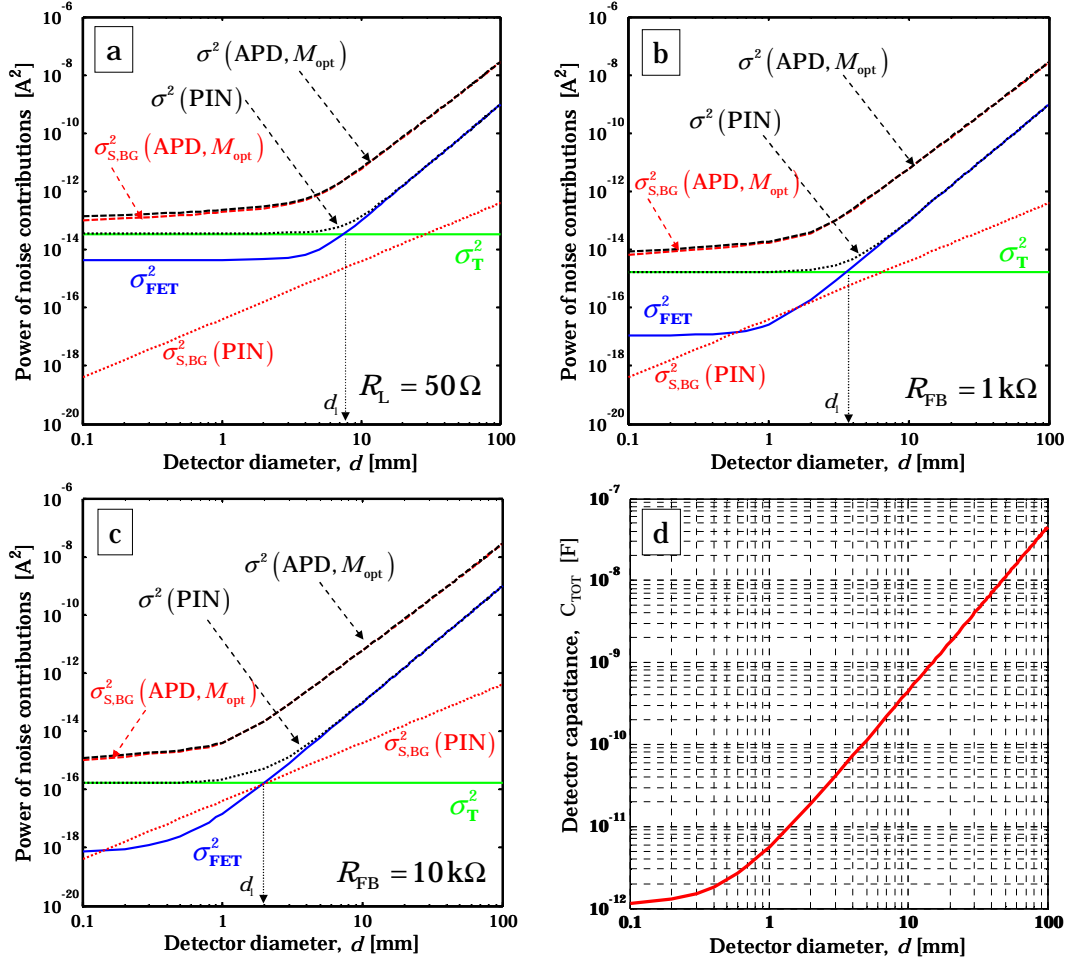


Figure 4.16: Power of the noise contributions for PIN and APD receivers in combination with a) low-impedance, b) and c) TIA preamplifier. d) Detector capacitance. Parameters are given in Table 4.2.

diameter $d < d_0$ incurs a slower increase of $\sigma_{\text{S,BG}}^2(\text{APD}, M_{\text{opt}})$ than $\sim d^2$, whereas the growth of M_{opt} for $d > d_0$ leads to an increase greater than $\sim d^2$, but smaller than $\sim d^4$. It can also be concluded that a larger feedback resistor in the TIA amplifier of an APD detector reduces the total noise for small detectors, but the maximum detector size for which the noise is well suppressed also decreases (i.e., the break point of the two runs of $\sigma_{\text{S,BG}}^2(\text{APD}, M_{\text{opt}})$ -curve shifts towards smaller d).

Finally, in Fig. 4.16 d, the increase of detector capacitance (C_{TOT} from (4.27)) with the diameter d is presented. It can be seen that, for detector diameters larger than approximately 1 mm, $C_{\text{det}} \sim d^2$ determines the run of C_{TOT} . As a consequence, when the capacitive component of the FET noise becomes dominant, this determines the run $\sim d^2$ of σ_{FET}^2 .

Figure 4.17 shows a comparison of the sensitivity for considered solutions in terms of minimum required received optical power $P_{\text{sig,min}}$ (see (4.42)). In order to make a comparison independent of SNR, all values are normalized with the optical power required in the case of a low-impedance 3 mm large PIN detector (point X in Fig. 4.17). Since there are commercially available detectors with diameters of some mm, this is the region of interest.

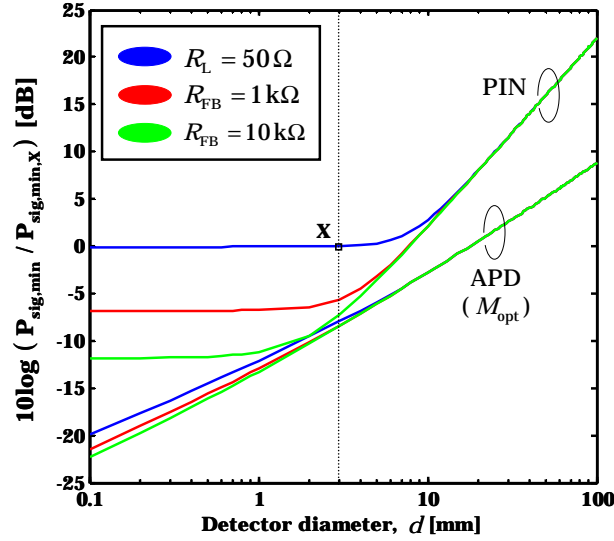


Figure 4.17: Comparison of sensitivity for different receivers.

If only the PIN receivers are compared, it can be observed again that a large feedback resistor ensures great (here, up to more than tenfold) enhancement in receiver sensitivity for small detectors. Nevertheless, when the FET capacitance becomes the dominant noise factor all receivers perform the same - the power requirement grows as $\sim d^2$. The same is true for the APD receivers, however, the enhancement in sensitivity even for small detectors is rather small because of the shot noise dominance. This is the reason why all the APD curves have approximately the same slope over the whole d -range ($\sqrt{F(M_{\text{opt}})}$ changes very slow with d , so the slope of the APD plots $\sim d$). With the large resistor (i.e., with the TIA preamplifier), the corresponding curves for PIN and APD run closer to each other. The minimum difference occurs at d_0 , where M_{opt} is closest to 1 (for the given load resistor).

Given that the same performance of PIN and APD would require a huge R_{FB} and a relatively small $d < 1$ mm (see Fig. 4.15), in this work an APD of $d = 3$ mm is chosen (with $C_{\text{TOT}}(d=3 \text{ mm}) \approx 40$ pF, see Fig. 4.16 d), together with a TIA preamplifier with a typical $R_{\text{FB}} = 10$ k Ω . Beside the advantage in sensitivity (even though not a great one) over a PIN with the same TIA, such receiver is shot-noise limited and the link analysis can be performed with the assumption of a white

Gaussian noise in the channel (as is usual for optical wireless links). On the other side, a corresponding PIN receiver is dominated by the FET noise component (see Fig. 4.16 c), which is frequency dependent. Moreover, the performance of an APD receiver is not as strongly dependent on the particular preamplifier design as the performance of a PIN one.

The bandwidth achievable with the regarded receiver depends, of course, on the preamplifier type (influencing the RC bandwidth) as well as the photodiode type (influencing the transit time bandwidth), which is already chosen to be an APD. The TIA preamplifier is much more convenient since the desired bandwidth can be set by the proper choice not only of the R_{FB} , but also of the open loop gain G , which can be manipulated by the choice of the amplifying transistor drain resistor [41]. This presents an additional degree of freedom compared to the case of the low impedance receiver.

Figure 4.18 shows the bandwidth of an APD receiver, where f_{RC} , f_{tr} and f_M components are shown along with the total receiver bandwidth f_g over the diameter size. Bandwidth f_M is calculated with a rule-of-thumb typical value for the Mf_M product of 200 GHz, and always with a corresponding M_{opt} . In Fig. 4.18, a low impedance preamplifier and a TIA preamplifier with $R_{FB} = 10 \text{ k}\Omega$ and $G = 250$ are compared.

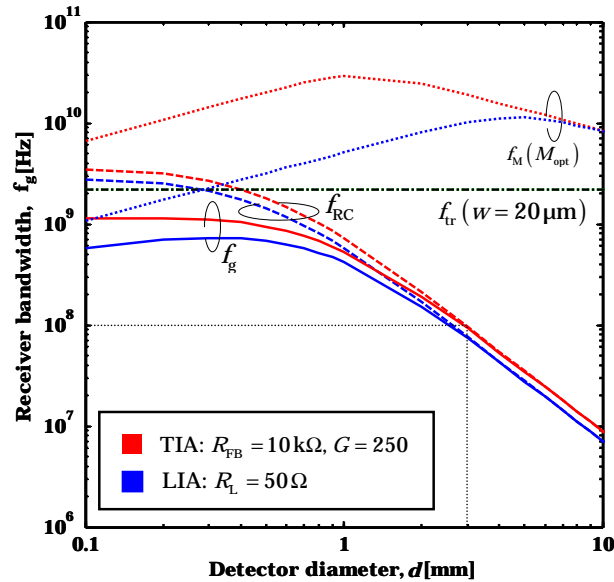


Figure 4.18: Bandwidth of an APD receiver with M_{opt} , for low impedance (dashed lines) and transimpedance (solid lines) preamplifier design.

It can be seen that for the small detector sizes, large optimal multiplication factors make the f_M bandwidth quite critical for the low impedance preamplifier, whereas

f_{tr} reduces the total bandwidth of the TIA preamplifier. In the case of large detectors (as well as in the region of interest, i.e., $d > 1$ mm), the RC -bandwidth is the dominant factor for both preamplifier types. In both cases shown, the desired bandwidth of 100 MHz can be realized with a detector whose diameter is about 3 mm (the difference is very small between the chosen TIA and low impedance preamplifier). However, note that the TIA curves can be moved to the right by choosing a larger G (if practically possible).

4.2.3 Summary of chosen parameters

The discussion in this chapter was not intended to optimize the receiver design as was done, e.g., in [41], but had the goal to provide a verified set of receiver parameters useful for further investigations, and yet realizable with realistic elements. Table 4.3 summarizes the OW receiver elements and chosen parameters.

Element	Parameter	Value (Quality)
Non-imaging Concentrator	type	hemispherical
	radius	$R_C = 5$ mm
	glass refraction index	$n = 1.8$
	field-of-view	$\Psi_C = \pi$ rad
Opt. filter	type	thin-film, hemispherical
	optical bandwidth	$\Delta\lambda = 15$ nm ¹⁹
	transmission coefficient	$T_0 = 0.8$
Photodetector	type	Si-APD
	multiplication factor	$M = 10.4$
	responsivity	$\eta = 0.7$ A/W
	ionization coefficient ratio	$k_A = 0.03$
	depletion region width	$w = 20$ μ m
	diameter	$d = 3$ mm
	input capacitance	$C_{TOT} \approx 40$ pF
Preamplifier	electrical RC -bandwidth	$f_g \approx f_{RC} \approx B = 100$ MHz
	type	TIA
	feed-back resistor	$R_{FB} = 10$ k Ω
	open-loop gain	$G = 250$

Table 4.3: Summary of the relevant OW receiver parameters for the target system.

An APD photodiode with diameter $d = 3$ mm ($C_{TOT}(d=3 \text{ mm}) \approx 40$ pF) and $M = 10.4$ is chosen in combination with a TIA preamplifier with a feed-back resistor $R_{FB} = 10$ k Ω and open loop amplification $G = 250$. In this way, a signal of 100 MHz bandwidth in the LOS link can be detected without equalization. Also, the noise components (when calculated with parameters from Table 4.3, and other

parameters for FET noise from Table 4.2) are $\sigma_S^2 \approx 8.3 \cdot 10^{-14} [\text{A}^2]$, $\sigma_T^2 \approx 1.6 \cdot 10^{-16} [\text{A}^2]$ and $\sigma_{\text{FET}}^2 \approx 8.14 \cdot 10^{-16} [\text{A}^2]$. Since $\sigma_S^2 \gg \sigma_T^2 + \sigma_{\text{FET}}^2$ (more than 10 times, which can also be seen in Fig. 4.16 c by $d = 3 \text{ mm}$), the SNR can be further on calculated by

$$\text{SNR} = \frac{\eta^2 M^2 P_{\text{sig}}^2}{2pqM^2 F(M) I_{\text{BG}} B}. \quad (4.47)$$

To provide an extra margin, which could account for greater power of background light, deployment of the receiver with $M \neq M_{\text{opt}}$ or similar occurrences, noise is additionally enlarged by 40% (factor p in (4.47)) for all the upcoming calculations.

Chapter 5

Evaluation of System Capacity

This chapter handles the evaluation of system capacity, according to the principles of information theory. In Section 5.1, capacity is defined, and the well-known expressions for capacity of flat (narrowband) and frequency-selective (wideband) power- and band-limited Gaussian channels are presented. The latter is the capacity of a conventional OFDM system in a frequency-selective channel. This capacity is obtained by the well-known water-filling algorithm. In this section, constraints on the input signal of the OW system are introduced. Since the capacity analysis of RF OFDM systems can not in the same way solve the problem for an OW OFDM system, Section 5.2 evaluates the capacity under these OW-specific constraints. In [49], You and Kahn developed a trigonometric moment space method to derive an upper-bound for the capacity of optical IM/DD channels with complex multiple subcarrier modulation and fixed bias. The result is valid for a flat optical channel with high signal-to-noise ratio (SNR). By generalization of this result, in this chapter, the trigonometric moment space method is used to provide an upper capacity bound of the system in a frequency-selective OW channel. Separately, the capacity upper bound according to another, conservative constraint is derived. Such bound is of relevance for one aspect of the investigation concerning the practical system which is regarded in Chapter 6. Given that the water-filling algorithm (as previously mentioned suitable for power-limited systems) is well-known, it can also be occasionally encountered in the literature as an approximative method for capacity evaluation of an optical IM/DD system (e.g., [91]). In Section 5.3, we explain the compromises and relaxations of the initial capacity problem which are needed for proper application of water-filling in an OW system. The resultant estimation of system capacity is then discussed.

5.1 Definitions and constraints

5.1.1 Capacity of the narrowband power- and band-limited Gaussian channel

Due to the nature of noise in it, OW channel is well modeled as Gaussian. Gaussian channel is a time discrete channel, where the output Y is given by the sum of mutually independent input X and Gaussian distributed noise¹ with a variance σ^2

$$Y = X + Z, \quad Z \sim \mathcal{N}(0, \sigma^2) \quad (5.1)$$

at any moment in time. Information theory defines the capacity of Gaussian channel as the maximum of the mutual information I between the channel input and output, over all possible input distributions $X \sim p(x)$ [80]

$$C_D = \max_{p(x)} I(X; Y), \quad (5.2)$$

which can be translated into a maximum difference of differential entropies h of the output and noise distributions

$$C_D = \max_{p(x)} (h(Y) - h(Z)). \quad (5.3)$$

If X can take any real value (unconstrained input), the capacity is infinite.² Nevertheless, this is in practice hardly ever the case. The most common constraint, encountered in practical systems, limits the second moment (power) of the input codeword (signal)

$$\mathbb{E}\{X^2\} \leq P. \quad (5.4)$$

In his famous work [81], Shannon has shown that in a Gaussian channel, power-limited with (5.4), a Gaussian distributed input $X \sim \mathcal{N}(0, P)$ maximizes the entropy (of the input and then indirectly of the output signal). Such input, therefore, achieves the capacity of the discrete time Gaussian channel as

$$C_D = \frac{1}{2} \log_2 \left(1 + \frac{P}{\sigma^2} \right) \quad [\text{bits/transmission/dim}], \quad (5.5)$$

where dim denotes dimension.³

¹Notation $\mathcal{N}(a, b)$ denotes Gaussian (normal) distribution with mean value a and variance b .

²An infinite subset of inputs can be chosen arbitrarily far apart, so that they are distinguishable at the output with arbitrarily small probability of error, regardless of noise.

³For example, transmission over a real-valued channel is one-dimensional, and over a complex-valued channel two-dimensional. Hence, the dimensions correspond to the degrees of freedom.

Practical channels are, however, usually limited in bandwidth and continuous in time (i.e., with continuous signals at the input and output). The Nyquist-Shannon sampling theorem [80] states that a function limited to bandwidth B can be entirely represented with $2B$ samples per second ($T_{\text{sam}} = 1/(2B)$). Also, even though a band-limited function is infinite in time, it is assumed that most of the energy is present in a finite time interval of length T_{sym} (defining an *almost* band-limited *almost* time-limited function). Based on the result for the discrete time channel (5.5), this led to the most popular definition for capacity of a power- and band-limited continuous channel:

The information capacity of a continuous channel of bandwidth B Hertz, perturbed by additive white Gaussian noise of two-sided power spectral density $N_0/2$ and limited in bandwidth to B , is given by

$$C = B \log_2 \left(1 + \frac{P}{N_0 B} \right) \quad [\text{bit/s/dim}], \quad (5.6)$$

and is achieved for Gaussian distributed input signal $\mathcal{N}(0, P)$.

Note that $N_0 B = \sigma^2$ from (5.5), which denotes the noise power per transmission, i.e., per channel use.

5.1.2 Capacity of the wideband (frequency selective) power- and band-limited Gaussian channel

With knowledge of (5.6) this subsection introduces the capacity of a frequency-selective, power- (P) and band-limited (B) Gaussian channel, according to the information theory. This result will serve as a basis for discussion regarding the capacity of the frequency selective optical wireless channel (e.g., see (3.19)).

A case of time-invariant frequency-selective channel is considered, whose frequency characteristic $H(f)$ (sketched in Fig. 5.1 a) is known both to the transmitter and receiver (ideal case). It is further assumed that frequency axis is divided into N subchannels of bandwidth $B_{\text{SC}} = B/N$, with constant (flat) fading, i.e., $H(f_n) = H_n$ in a subchannel n . A frequency selective channel can thus be represented over a set of N parallel (independent) AWGN channels

$$Y_n = H_n X_n + Z_n, \quad n = 1, \dots, N. \quad (5.7)$$

Its capacity is given as the sum of capacities of parallel Gaussian channels, with input signals constrained by a *common* power limit

$$\mathbb{E} \left\{ \sum X_n^2 \right\} = \sum \mathbb{E} \{ X_n^2 \} = \sum P_n \leq P, \quad (5.8)$$

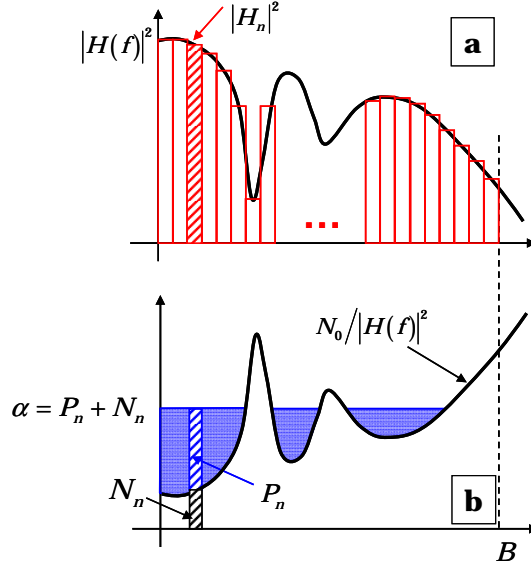


Figure 5.1: (a) A frequency response of a wideband channel seen as parallel flat channels (given merely as a sketch), (b) Illustration of the “water-filling” algorithm - sum of signal power and effective noise power (including the channel influence) should reach to a constant level).

where $P_n = \mathbb{E}\{X_n^2\} = \langle X_n^2 \rangle$ is the signal power (second moment) in the n^{th} sub-channel. The capacity is therefore given as

$$C = \max_{\sum P_n \leq P} B_{\text{SC}} \sum_{n=1}^N \log_2 \left(1 + \frac{P_n |H_n|^2}{N_0 B_{\text{SC}}} \right) \quad [\text{bits/s/dim}], \quad (5.9)$$

where N_0 denotes the (single-sided) power spectral density of the white Gaussian noise.

Capacity is reached when the input powers P_n are optimally chosen. The optimal power allocation is determined by the *Lagrange multiplier method* [80], [82], which renders extrema of constrained functions. This method finds a vector of input powers $[P_1, \dots, P_N]$ which maximizes the N -variable target function (5.9), under the constraint $g(P_1, \dots, P_N) = \sum P_n - P = 0$. In that context, a Lagrange function is formed as

$$L(P_1, \dots, P_N) = C(P_1, \dots, P_N) - \alpha g(P_1, \dots, P_N) \quad (5.10)$$

where a real-valued parameter $\alpha \in \mathbb{R}$ (called Lagrange multiplier) is introduced for each existing constraint (in this case, only one). The optimization solution is then obtained from the system of N equations

$$\frac{\partial L(P_1, \dots, P_N)}{\partial P_n} = 0 \quad n = 1, \dots, N. \quad (5.11)$$

In this way, a constrained problem is converted into an unconstrained one. The optimal power allocation is found as

$$P_n = \begin{cases} \alpha - N_n, & N_n < \alpha, \\ 0, & \text{otherwise,} \end{cases} \quad (5.12)$$

where

$$N_n = \frac{N_0}{|H_n|^2} B_{\text{SC}} \quad (5.13)$$

is the effective noise power in subchannel n , including the channel influence (black shaded bars in Fig. 5.1 a). Optimization result (5.12) means that the function (5.9) is maximized if the signal power and the effective noise power add up to a constant level in each subchannel. Illustratively, this optimal power allocation resembles a process of filling a vessel with an uneven bottom with an amount of water P (as in Fig. 5.1 b), and it is therefore in the communications community known as the “water-filling” algorithm.

5.1.3 Constraints on the transmit signal in OW systems

Even though the equivalent baseband model of an OW multipath link mathematically looks the same as of radio or wired electrical systems (LTI system, introduced in (2.5)), due to IM/DD transmission on one side, and safety regulations on the other, rather than on signal power (second moment) there are limits on the first moment and on non-negativity of the input signal. If the mean optical power at the source is limited to P_O , constraints on the signal x in front of the optical source (in time domain, introduced by (3.24)) are

$$(E\{x\} \leq P_O/S) \quad \wedge \quad (x \geq 0), \quad (5.14)$$

where S [W/A] denotes the source conversion factor.

From (5.14), it can easily be concluded that the Gaussian distribution at the input of the OW channel cannot be capacity-achieving. Namely, when the mean value is finite, meeting the non-negativity constraint with a Gaussian distribution forces the signal variance (AC power) towards zero. However, not only that the ultimate capacity of the OW channel is not given by the Shannon formula (5.5), but it seems that an analogous closed-form expression is hardly obtainable. This is primarily because the capacity-achieving input distribution can vary over the SNR range. Moreover, such distributions can be often analytically derived only under some extreme conditions, leading not to true capacity, but to its bounds, which are more or less tight. For example, Shannon found in [81] that under conditions (5.14) on a random input $X \sim p(x)$, the maximum of differential entropy $\max h(X)$ is obtained with the exponential distribution

$$p(x) = \begin{cases} (S/P_O)e^{-xS/P_O}, & x > 0, \\ 0, & x \leq 0, \end{cases} \quad (5.15)$$

However, since capacity is achieved for $\max h(Y) = \max h(X + Z)$, the expression for capacity, obtained with exponential distribution at the input only asymptotically approaches the true capacity for sufficiently high SNR.

In this chapter, the analysis of the capacity in multiple-subcarrier (OFDM) based OW systems is regarded. As shall be seen, the exponential distribution is not capacity-achieving in this case (not even for a bound), because the sinusoidal waveforms of the subcarriers in time domain affect the choice of input distribution. The same is valid for distributions which were proposed in [83] as optimal for capacity bounds in the base-band IM/DD channels. The goal of this chapter is to discuss the capacity of a system described in Chapter 3.4, where the signal (3.23) complies to the set of constraints (5.14).

5.2 TMS method for MSM-based OW systems

In an often cited work [49], You and Kahn derived an upper bound for capacity of optical IM/DD channels with complex multiple subcarrier modulation (MSM) and fixed bias, using trigonometric moment space (TMS) method. Their result is valid for a flat optical channel with high SNR. This method can be used to upper-bound the capacity of the system regarded in this thesis. Nevertheless, since the considered channel varies with system geometry, some generalization of the result from [49] is necessary.

In Section 5.2.1, the main steps of TMS method are revised. In Section 5.2.2, the derivation of the capacity upper bound is performed directly for the frequency-selective channel. Moreover, the derivation is based on the OFDM system presented in the Chapter 3.4 (digital realization), unlike in [49] where analogue filter banks were considered. Even though this does not influence the final result, it was chosen for the sake of consistency.

5.2.1 Trigonometric moment space method

The received complex symbol Y_n on the n^{th} ($n = 1, \dots, N - 1$) subcarrier of the OFDM system from Chapter 3 (3.19) can be represented as

$$Y_n = H_{eq}(f_n) X_n + Z_n, \quad (5.16)$$

where $X_n = a_n + jb_n$ is the sent complex symbol (signal constellation) at frequency f_n , and $Z_n = Z_{I,n} + jZ_{Q,n}$ is the complex AWGN. Noise components, $Z_{I,n}$ and $Z_{Q,n}$, are independent identically distributed (i.i.d.) with $\mathcal{N}(0, N_{\text{SC}})$, where variance N_{SC} in $[\text{A}^2]$ is given by

$$N_{\text{SC}} = \text{E}\{Z_{I,n}^2\} = \text{E}\{Z_{Q,n}^2\} = \frac{1}{2} \frac{N_0}{2} B_{\text{SC}} (2N)^2 = (2N)^2 \frac{N_0}{4} B_{\text{SC}}. \quad (5.17)$$

Hereby, $N_0/2$ is the (two-sided) noise power spectral density directly after the photodiode, and the scaling factor $(2N)^2$ comes from the chosen mathematical definition of the FFT operator ((3.8) and (B.2)). Bandwidth of one subcarrier is given by $B_{SC} = B/N$. As introduced in Chapter 3.4, the equivalent channel coefficient at a particular frequency is given by the corresponding coefficient of the optical wireless channel scaled with the receiver responsivity η and conversion factor of the optical source S

$$H_{eq}(f_n) = \eta H(f_n) S. \quad (5.18)$$

The conversion factor at the optical source can be set to $S = 1 \text{ W/A}$ without loss of generality, since in fact it does not influence the channel capacity (in case of the mean optical power constraint), as shall be seen later on.

Transmission on all independent subcarriers (in a system illustrated in Fig. 3.6 and Fig. 3.8) can be encompassed by a matrix-vector notation as

$$\begin{aligned} [\mathbf{Y}] &= [\mathbf{H}_{eq}][\mathbf{X}] + [\mathbf{Z}] \\ [Y_n]_{n=1}^{N-1} &= [\mathbf{H}_{eq}][X_n]_{n=1}^{N-1} + [Z_n]_{n=1}^{N-1}. \end{aligned} \quad (5.19)$$

Notation $[\mathbf{Y}] = [Y_1 \dots Y_{N-1}]^T = [Y_n]_{n=1}^{N-1}$ is adopted for vectors in the second equation. Since the subcarrier frequencies are mutually orthogonal (and $[\mathbf{H}_{eq}]$ is diagonal), transmission takes place over parallel Gaussian channels. Further on, since symbols on the subcarriers are complex, the system with $N - 1$ independent subcarriers is equivalent to $2(N - 1)$ parallel Gaussian channels (i.e., for each X_n , a_n and b_n are independent). Note that, because each of these subchannels is flat, and the noise is independent identically distributed from symbol to symbol, the capacity can be achieved with independent identically distributed inputs and it is sufficient to consider only one signaling interval.

A goal is, therefore, to characterize the allowed input vectors and to describe the sequences which satisfy the constraints (5.14) on the signal in time domain (given by (3.23), i.e., (3.24)). The mean power constraint can be immediately incorporated. Since

$$E\{x\} = \frac{1}{T_{\text{sym}}} \int_0^{T_{\text{sym}}} x(t) dt = \frac{1}{2N} X_0 \leq \frac{P_O}{S}, \quad (5.20)$$

with $X_0 = 2NP_O/S$ the required condition is satisfied. If we introduce a base angular frequency $\omega_0 = 2\pi/T_{\text{sym}}$, the transmission signal (3.23) can be brought to a more convenient form

$$\begin{aligned} x(t) &= \frac{P_O}{S} + \frac{P_O}{S} \sum_{n=1}^{N-1} \left(\frac{2S}{2NP_O} a_n \cos(n\omega_0 t) - \frac{2S}{2NP_O} b_n \sin(n\omega_0 t) \right) \\ &= \frac{P_O}{S} \left(1 + \sum_{n=1}^{N-1} (a_n^1 \cos(n\omega_0 t) - b_n^1 \sin(n\omega_0 t)) \right), \end{aligned} \quad (5.21)$$

where $a_n^1 = 2Sa_n/(2NP_O)$ and $b_n^1 = 2Sb_n/(2NP_O)$. If an auxiliary sequence $[X_n^1]_{n=1}^{N-1}$ is defined as $X_n^1 = a_n^1/2 + jb_n^1/2$, the aim is reformulated to finding the necessary and sufficient conditions for this sequence to fulfill $x(t) \geq 0$. As in [49], there is an $(N-1)$ -dimensional space \mathcal{CH}_{N-1} , defined by all the sequences $[X_n^1]_{n=1}^{N-1}$ for which the needed condition is met

$$\mathcal{CH}_{N-1} = \left\{ \begin{array}{l} [X_n^1 = a_n^1/2 + jb_n^1/2]_{n=1}^{N-1} : \\ 1 + \sum_{n=1}^{N-1} (a_n^1 \cos(n\omega_0 t) - b_n^1 \sin(n\omega_0 t)) \geq 0 \end{array} \right\}. \quad (5.22)$$

In [49], it was shown that this problem can not be exactly solved for a band-limited function $x(t)$. The problem was, however, solved for a function $x'(t)$ with a lifted bandwidth limit. It has been shown that such relaxation leads to the upper bound on channel capacity.

It is known that Fourier coefficients (which are infinitely long sequences) of non-negative functions (Fourier series) lie inside a convex trigonometric moment space (TMS) \mathcal{M} with infinite number of dimensions [49]. The truncation of the sequence after the first $N-1$ terms must fall into a finite-dimensional trigonometric moment space

$$\mathcal{M}_{N-1} = \left\{ \begin{array}{l} [X_n^1 = a_n^1/2 + jb_n^1/2]_{n=1}^{N-1} : \\ X_n^1 = \frac{1}{T_{\text{sym}}} \int_0^{T_{\text{sym}}} x'(t) e^{-jn\omega_0 t} dt, \quad \frac{1}{T_{\text{sym}}} \int_0^{T_{\text{sym}}} x'(t) dt = 1 \end{array} \right\}. \quad (5.23)$$

It has been proven that \mathcal{CH}_{N-1} is a subset of \mathcal{M}_{N-1} , i.e., $\mathcal{CH}_{N-1} \subseteq \mathcal{M}_{N-1}$.

Geometrical characterization of the space \mathcal{M}_{N-1} , more precisely the determination of its volume is the main step in derivation of the upper capacity bound. It has first been shown that the finite trigonometric moment space of interest is equivalent to the *convex hull*⁴ of the space curve

$$\mathcal{E} = \{[e^{-jn\omega_0 t}]_{n=1}^{N-1}, \quad t \in [0, 2\pi/\omega_0]\}, \quad (5.24)$$

$$\mathcal{M}_{N-1} = \text{conv}(\mathcal{E}).$$

An equivalent representation of \mathcal{E} in a $2(N-1)$ -dimensional real-valued space is

$$\mathcal{E} = \{[\cos(\omega_0 t) \quad -\sin(\omega_0 t) \quad \cos(2\omega_0 t) \quad -\sin(2\omega_0 t) \quad \dots \quad \cos((N-1)\omega_0 t) \quad -\sin((N-1)\omega_0 t)], \quad t \in [0, 2\pi/\omega_0]\}. \quad (5.25)$$

The volume of \mathcal{M}_{N-1} was obtained by recognizing that \mathcal{E} presents a linear transformation of the generalized ellipse. A generalized ellipse space curve, described with

$$\mathcal{GE} = \left\{ \left[\frac{1}{n} (\cos(n\omega_0 t) + j \sin(n\omega_0 t)) \right]_{n=1}^{N-1}, \quad t \in \left[0, \frac{2\pi}{\omega_0} \right] \right\}, \quad (5.26)$$

⁴The convex hull of a set is defined as the smallest convex set containing the original set [84].

can be represented also by

$$\mathcal{GE} = \left\{ \begin{bmatrix} \cos(\omega_0 t) & \sin(\omega_0 t) & \frac{1}{2} \cos(2\omega_0 t) & \frac{1}{2} \sin(2\omega_0 t) \cdots \\ \cdots & \frac{1}{N-1} \cos((N-1)\omega_0 t) & \frac{1}{N-1} \sin((N-1)\omega_0 t) \end{bmatrix}, \quad t \in \left[0, \frac{2\pi}{\omega_0}\right] \right\}. \quad (5.27)$$

Since $\mathcal{E} = [\mathbf{A}] \cdot \mathcal{GE}$, where $[\mathbf{A}]$ is a diagonal matrix

$$[\mathbf{A}] = \mathbf{diag}[1 \quad -1 \quad 2 \quad -2 \cdots (N-1) \quad -(N-1)], \quad (5.28)$$

the volume of the desired TMS can be derived from the volume of the convex hull of generalized ellipse as [85]

$$\begin{aligned} V_{\mathcal{M}_{N-1}} &= \det([\mathbf{A}]) V_{\text{conv}(\mathcal{GE})} \\ &= [(N-1)!]^2 \frac{(2\pi)^{N-1}}{(N-1)!(2N-2)!} = \frac{(N-1)!(2\pi)^{N-1}}{(2N-2)!}, \end{aligned} \quad (5.29)$$

where $\det(\cdot)$ denotes matrix determinant, and $V_{\text{conv}(\mathcal{GE})}$ was calculated in [86].

5.2.2 Derivation of the capacity upper bound

As already mentioned, transmission of independent complex-valued symbols over $N-1$ orthogonal subcarriers is equivalent to the transmission over $2(N-1)$ parallel (real-valued) Gaussian channels. The input-output relationship between the two frequency domains in an OFDM system can be represented by a matrix equation (5.19), where $(N-1)$ -long vectors have complex-valued elements, or equivalently with $2(N-1)$ -long vectors of real valued elements. So the channel input $[\mathbf{X}]$ and output $[\mathbf{Y}]$ can be written as

$$[\mathbf{X}]_{2(N-1) \times 1} = \begin{bmatrix} a_n \\ b_n \end{bmatrix}_{n=1}^{N-1} \quad (5.30)$$

$$[\mathbf{Y}]_{2(N-1) \times 1} = \begin{bmatrix} \hat{a}_n \\ \hat{b}_n \end{bmatrix}_{n=1}^{N-1} = \begin{bmatrix} \eta H_n S a_n + Z_{I,n} \\ \eta H_n S b_n + Z_{Q,n} \end{bmatrix}_{n=1}^{N-1}. \quad (5.31)$$

The upper bound of discrete system capacity in [bits/transmission] is then calculated according to the definition (5.2), so the upper bound of C in terms of theoretically possible transmission rate in [bits/s] is given by

$$C = \frac{1}{T_{\text{sym}}} \max_{\mathbf{X} \in \mathcal{M}_{N-1}} (h([\mathbf{Y}]) - h([\mathbf{Z}])), \quad (5.32)$$

where T_{sym} is the duration of one transmission, i.e., an OFDM symbol period.

Differential entropy of noise can be easily obtained. $[\mathbf{Z}]$ has a Gaussian distribution with variance N_{SC} in each of $2(N-1)$ channels, and the entropy of the multivariate normal distribution $[\mathbf{Z}] \sim \mathcal{N}(0, [\mathbf{K}])$ with covariance matrix $[\mathbf{K}]$ is given by [80]

$$h([\mathbf{Z}]) = \frac{1}{2} \log_2 (2\pi e)^{2(N-1)} \det([\mathbf{K}]). \quad (5.33)$$

In the case of orthogonal (i.e., independent) channels, $[\mathbf{K}]$ is a diagonal matrix with variance of each channel on the diagonal. Hence,

$$h([\mathbf{Z}]) = \frac{1}{2} \log_2 (2\pi e N_{\text{SC}})^{2(N-1)}. \quad (5.34)$$

For high SNRs, $h([\mathbf{Y}])$ can be calculated based on the volume of TMS from the previous subsection (5.29). When $[\mathbf{H}_{\text{eq}}][\mathbf{X}] \gg [\mathbf{Z}]$, $[\mathbf{Y}]$ is almost entirely bounded within the signal space where $[\mathbf{H}_{\text{eq}}][\mathbf{X}]$ is distributed, and its entropy mostly consists of the contribution of $h([\mathbf{H}_{\text{eq}}][\mathbf{X}])$ with some residue. Therefore,

$$\begin{aligned} h([\mathbf{Y}]) &\approx h \left(\begin{bmatrix} \eta H_1 S a_1 \\ \eta H_1 S b_1 \\ \vdots \\ \eta H_{N-1} S a_{N-1} \\ \eta H_{N-1} S b_{N-1} \end{bmatrix} \right) = h \left(\begin{bmatrix} \eta H_1 (2NP_O)(a_1^1/2) \\ \eta H_1 (2NP_O)(b_1^1/2) \\ \vdots \\ \eta H_{N-1} (2NP_O)(a_{N-1}^1/2) \\ \eta H_{N-1} (2NP_O)(b_{N-1}^1/2) \end{bmatrix} \right) \\ &= \log_2 \left((2N\eta P_O)^{2(N-1)} \prod_{n=1}^{N-1} |H_n|^2 \right) + h \left([a_1^1/2 \ b_1^1/2 \cdots a_{N-1}^1/2 \ b_{N-1}^1/2]^T \right) \\ &= \log_2 \left((2N\eta P_O)^{2(N-1)} \prod_{n=1}^{N-1} |H_n|^2 \right) + h([\mathbf{X}^1]). \end{aligned} \quad (5.35)$$

In (5.35), vector $[\mathbf{X}^1] = [a_1^1/2 \cdots b_{N-1}^1/2]^T$ with elements $\Re(X_n^1)$ and $\Im(X_n^1)$, defined in (5.23), is explicitly extracted, since $[X_n^1]_{n=1}^{N-1}$ defines the sequence which falls within the TMS whose volume is given by (5.29). In [49], the authors proved that

$$h([\mathbf{Y}]) \leq \log_2 \left((2N\eta P_O)^{2(N-1)} \prod_{n=1}^{N-1} |H_n|^2 \right) + h([\mathbf{X}_{\text{uniform}}^1]) + o(\sigma), \quad (5.36)$$

where $[\mathbf{X}_{\text{uniform}}^1]$ denotes a vector with the uniform distribution within \mathcal{M}_{N-1} and where $o(\sigma)$ is the entropy residue, which disappears exponentially fast as $N_0 \rightarrow 0$ (i.e., at high SNRs). The differential entropy of a vector with uniform distribution within a certain space, is determined over the volume of this space

$$\begin{aligned} h([\mathbf{X}_{\text{uniform}}^1]) &= - \int_{\mathcal{M}_{N-1}} \frac{1}{V_{\mathcal{M}_{N-1}}} \log_2 \left(\frac{1}{V_{\mathcal{M}_{N-1}}} \right) dV_{\mathcal{M}_{N-1}} \\ &= \log_2 V_{\mathcal{M}_{N-1}}. \end{aligned} \quad (5.37)$$

Finally, introducing (5.36), (5.37) and (5.34) into (5.32), the capacity upper bound can be written as

$$\begin{aligned}
C &\approx \frac{1}{T_{sym}} \left\{ \log_2 V_{\mathcal{M}_{N-1}} \right. \\
&\quad \left. + \log_2 \left((2N\eta P_O)^{2(N-1)} \prod_{n=1}^{N-1} |H_n|^2 \right) - \frac{1}{2} \log_2 (2\pi e N_{SC})^{2(N-1)} \right\} \\
&= B_{SC} \left\{ \log_2 \frac{(N-1)!(2\pi)^{N-1}}{(2N-2)!} + (N-1) \log_2 \frac{\eta^2 P_O^2}{2\pi e \frac{N_0}{4} B_{SC}} + \log_2 \prod_{n=1}^{N-1} |H_n|^2 \right\} \\
&\approx (N-1) B_{SC} \log_2 \left(\frac{1}{4(N-1) 2^{\frac{1}{2(N-1)}} \frac{N_0}{4} B_{SC}} \frac{\eta^2 P_O^2}{N_0 B_{SC}} \right) + B_{SC} \log_2 \prod_{n=1}^{N-1} |H_n|^2 \\
&= B_{SC} \sum_{n=1}^{N-1} \log_2 \left(\frac{1}{(N-1) 2^{\frac{1}{2(N-1)}} \frac{N_0}{4} B_{SC}} \frac{\eta^2 P_O^2 |H_n|^2}{N_0 B_{SC}} \right) \\
&= \frac{B}{N} \sum_{n=1}^{N-1} \log_2 \left(\frac{N}{(N-1) 2^{\frac{1}{2(N-1)}} \frac{N_0}{4} B} \frac{\eta^2 P_O^2 |H_n|^2}{N_0 B} \right)
\end{aligned} \tag{5.38}$$

where Stirling approximation $x! \approx \sqrt{2\pi x} x^x e^{-x}$ is used for the factorial.⁵ Note that the actual SNRs (on each subcarrier) cannot be explicitly recognized, since the allowed ranges of the sequence in specific dimensions (ranges for a_n and b_n for any n) are incorporated in the volume of the trigonometric moment space. Hence, a new parameter $\gamma_n = 4\eta^2 P_O^2 |H_n|^2 / N_0 B_{SC}$ is chosen to be used as an SNR-analogue ($\gamma_n \neq \text{SNR}_n$) in following presentations of results.⁶

The capacity of a flat channel obtained in [49], presents a special case of (5.38) when $|H_n| = |H|$, for all n . Then

$$C = (N-1) B_{SC} \log_2 \left(\frac{1}{4(N-1) 2^{\frac{1}{2(N-1)}} \frac{N_0}{4} B_{SC}} \frac{\eta^2 P_O^2 |H|^2}{N_0 B_{SC}} \right). \tag{5.39}$$

In Fig. 5.2 a, normalized total capacity C/B_{SC} is shown depending on the γ -factor, in the case of a flat channel ($\gamma = \gamma_n$) and for a system with independent transmission over $N-1$ subcarriers. In this figure, bandwidth of one subcarrier is considered constant, so that the increase in the number of subcarriers actually results in an increase of the total transmission (channel) bandwidth $B = NB_{SC}$. (Investigations

⁵Due to the used approximation the obtained analytical expression does not render the tightest upper bound. A more exact upper bound could be obtained numerically, but the Stirling approximation gets better with larger x .

⁶ γ -factor presents a parameter analogous to the SNR on one subcarrier with noise in one dimension. In [49], this SNR-equivalent parameter is defined over the noise power in two dimensions. This means that the values of γ correspond to 3dB smaller values of SNR from [49].

when B is kept constant will follow.) It can be seen that, for a relatively good channel (i.e., large γ), higher capacity is obtained with more subcarriers. Larger N increases the total capacity, but the increment is not linear, because each subsequently added subcarrier is constrained by all the previous ones. Figure 5.2 b presents the average normalized capacity of one subcarrier $C/B_{SC}/(N - 1)$ over γ (since the exact normalized capacity can not be determined). With higher N , the average normalized capacity of one subcarrier is smaller (for a constant γ), which can be explained by the fact that the same power P_O is divided among more dimensions (and subcarriers - each subcarrier is equivalent to two dimensions in case of complex input signal constellations). For example, when N is doubled, the difference between the average normalized capacities pro subcarrier (5.39)

$$\begin{aligned} \frac{C_N}{(N-1)B_{SC}} - \frac{C_{2N}}{(2N-1)B_{SC}} &= 1 + \frac{1}{4N-2} - \frac{1}{2N-1} \\ &\rightarrow 1[\text{bit/s/Hz}] \quad \text{for } N \gg 1, \end{aligned} \quad (5.40)$$

which can be observed also in Fig. 5.2 b.

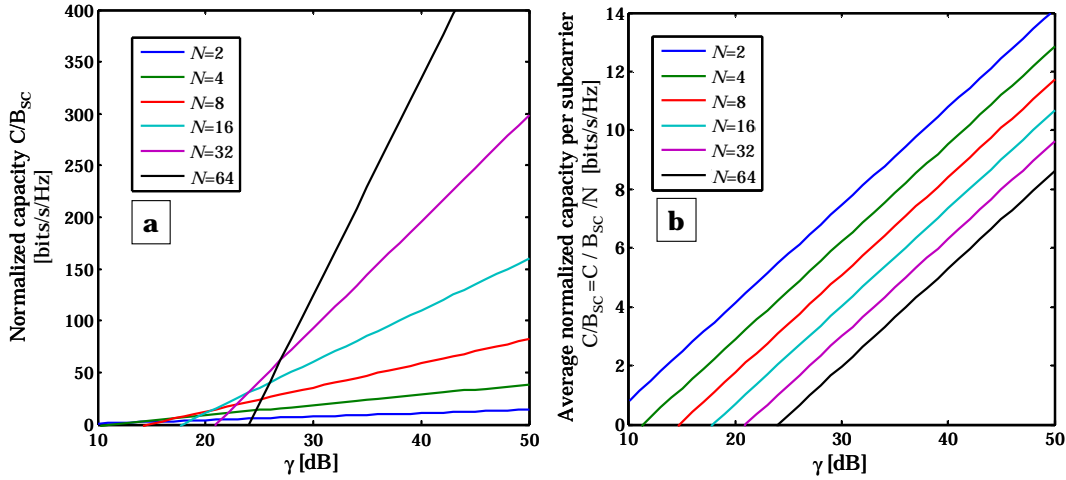


Figure 5.2: (a) Total normalized capacity in the flat channel ($\gamma = \gamma_n$) with $N - 1$ independent subcarriers, depending on the γ -factor, (b) Average normalized capacity per subcarrier.

From Fig. 5.2, it is also obvious that the method is not suitable (correct) for insufficiently high γ . That means that in a frequency selective channel, where γ is not constant across the bandwidth covered by the subcarriers, capacity can be evaluated by this method when the frequencies with small γ_n are not used for transmission. In (5.38), one recognizes the geometric mean of the channel gains,

so the third equality of (5.38) becomes

$$\begin{aligned} C &\approx (N-1)B_{\text{SC}}\log_2\left(\frac{1}{4(N-1)2^{\frac{1}{2(N-1)}}\frac{N_0}{4}B_{\text{SC}}}\frac{\eta^2 P_{\text{O}}^2}{N_0 B_{\text{SC}}}\right) + B_{\text{SC}}\log_2(|H|_{\text{g}}^2)^{N-1} \\ &= (N-1)B_{\text{SC}}\log_2\left(\frac{1}{4(N-1)2^{\frac{1}{2(N-1)}}}\gamma_g\right) \end{aligned} \quad (5.41)$$

where γ_g is the geometric mean of γ_n -factors over the used bandwidth

$$\gamma_g = \frac{4\eta^2 P_{\text{O}}^2}{N_0 B_{\text{SC}}} |H|_{\text{g}}^2 = \frac{4\eta^2 P_{\text{O}}^2}{N_0 B_{\text{SC}}} \sqrt[N-1]{\prod_{n=1}^{N-1} |H_n|^2}. \quad (5.42)$$

Since the capacity depends on the geometrical mean of the channel coefficients across the used bandwidth, in the case of a channel characteristic where subchannel coefficients strongly vary (or relatively large dips appear), it can be expected that it is not desirable to use the whole bandwidth, but only some part of it (in a way, similarly as for water-filling). This is explained by the following example.

Figure 5.3 a illustrates a channel characteristic of the optical wireless system (as introduced in Chapter 2 by (2.7), or similarly to Fig. 2.10), for several different K -factors (ratios of the LOS and diffuse channel component at the receiver) over the frequency band of $B = 100$ MHz. As the K -factor grows from -15 to +25 dB, the channel characteristic varies from low pass to almost flat (channel gains bandwidth and magnitude). A real-valued OFDM system is assumed with $N - 1 = 63$ independent subcarriers within this bandwidth. The x -axis shows the subcarrier indices (1 to 63, with adjacent spacing $B_{\text{SC}} = B/N = 100/64$ MHz). Figure 5.3 b illustrates the normalized capacity obtained for these channel characteristics, when the number of the used subcarriers grows (while $B_{\text{SC}} = \text{const.}$). Hereby, the optical power limit of $P_{\text{O}} = 400$ mW is assumed at the transmitter and bright sky as the noise source⁷. It can be observed that for low K -factors, using “too many” subcarriers diminishes the capacity. Hence, there is an optimal number of subcarriers, that is, an optimal portion of the bandwidth B recommendable for transmission. To evaluate the upper bound, the combination of subcarriers which renders the highest capacity should be chosen. This is the combination which offers the best trade-off between γ_g and the number of used subcarriers. For example, when $K = -5$ dB, the best would be to modulate some 30 subcarriers (about half of the 100 MHz bandwidth), whereas for $K = -15$ dB, the maximum would be achieved with about only 10 used subcarriers. Clearly, the optimal number of used subcarriers grows with the K -factor. Hence, in the case of relatively flat channel, there is no significant discrepancy among the channel gains over the whole bandwidth,

⁷The values for power limit and noise power assumed in calculations are consequence of investigations performed in Chapter 4, and can here be considered as illustrative.

and the normalized capacity curves are monotonically increasing with γ .

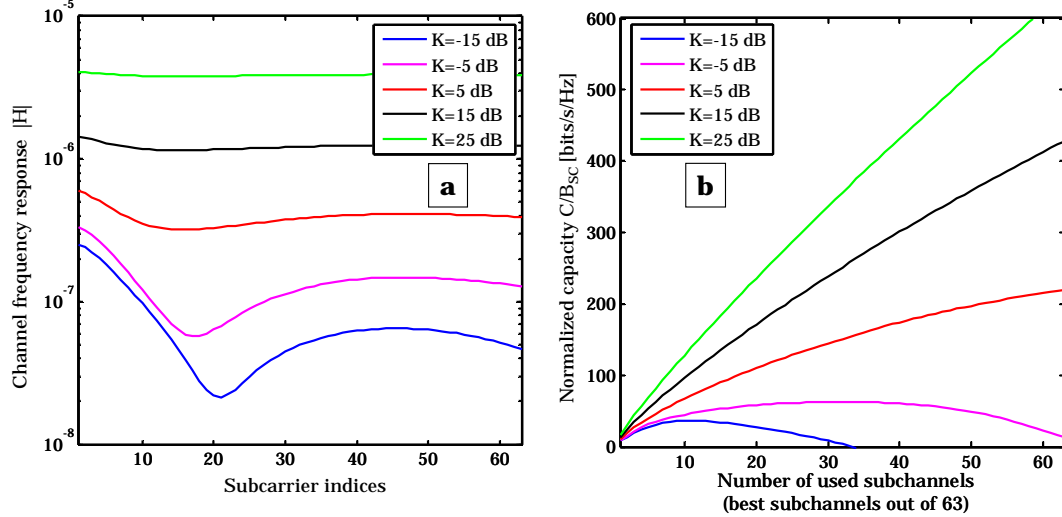


Figure 5.3: (a) Characteristic of a frequency-selective channel for different K -factors over the frequency axes, given in terms of subcarrier indices, (b) Total spectral efficiency, resulting from the trigonometric moment space method, for different number of used subcarriers and K -factor as parameter. Maximum bandwidth B and B_{sc} are constant; as “best” are denoted the subcarriers with lowest channel attenuation.

Figure 5.4 shows a plot of individual γ_n values calculated in all used subchannels (x -axis) and over a span of K -factors from -15 dB to $+25$ dB (y -axis). Four plots (Fig. 5.3 a-d) are presented for different optical power limits $P_O = 100, 200, 400$ and 800 mW, respectively, while the other system parameters are assumed the same as in previous calculations. For each K -factor, the optimal group of subcarriers is determined as previously described and their γ_n are plotted. Unused subcarriers are illustrated by dark blue color. In each of the figures, we can observe that with larger K more and more bandwidth is used and the values of γ_n generally grow. When comparing Fig. 5.4 c ($P_O = 400$ mW) to the channel characteristics in Fig. 5.3 a we can see that the unused subcarrier areas are there where the channel characteristic has dips. When looking at the variation of P_O , it is clear that with a better link budget (more power allowed in the system), the area of poor (unused) subcarriers diminishes and that the individual γ_n factors grow.

After choosing the optimal set of subcarriers ($N - 1$ becomes $N_{\text{used}} \leq N - 1$, while $B_{sc} = B/N$ remains constant), capacity is influenced by the product of the channel coefficients at these subcarrier frequencies. Therefore, the capacity is the same as of a flat channel (with the same bandwidth $(1 + N_{\text{used}})B_{sc}$), whose γ is equal to the geometric mean of the γ -factors of the used subcarriers in the frequency selective channel. Figure 5.5 shows the obtained values of γ_g (right y -axis) and the

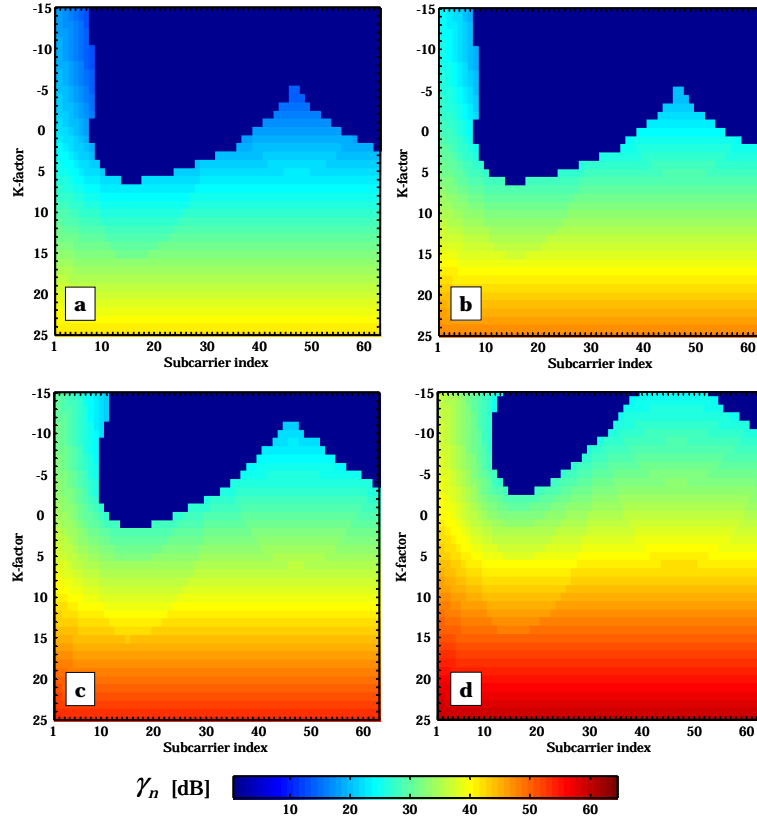


Figure 5.4: Plots of the γ -factor distribution across the subchannels and the foreseen K -factor range, for a mean optical power limit of (a) $P_O = 100$ mW, (b) $P_O = 200$ mW, (c) $P_O = 400$ mW, and (d) $P_O = 800$ mW.

corresponding number of used subcarriers (left y -axis) over different K -factors, for realistic parameters of the considered system. Two optical power limits are chosen: $P_O = 100$ mW (a) and 400 mW (b). When we translate the mutually corresponding (γ_g and N_{used}) results given in Fig. 5.5, to the Fig. 5.2 a, we see that the γ values (indirectly SNR), obtained for the considered system parameters fall in the area for which the upper bound derived by the trigonometric moment space method is applicable. Hence, the upper capacity bound for the frequency selective channel can be evaluated by (5.38), with the optimal set of subcarriers.

The result is presented in Fig. 5.6 in terms of transmission rate for different channel states (K -factor) and two optical power limits $P_O = 100$ mW and 400 mW. For comparison, curves for a *statically designed* multiple subcarrier system, aimed to guarantee the same transmission performance over the whole area of interest (in this work, the whole desk-top surface in the room model from Fig. 2.1) are also shown. In Fig. 5.6, these plots are referred to as non-adaptive solution. They are obtained also by (5.38), in the case of a purely diffuse channel (worst case

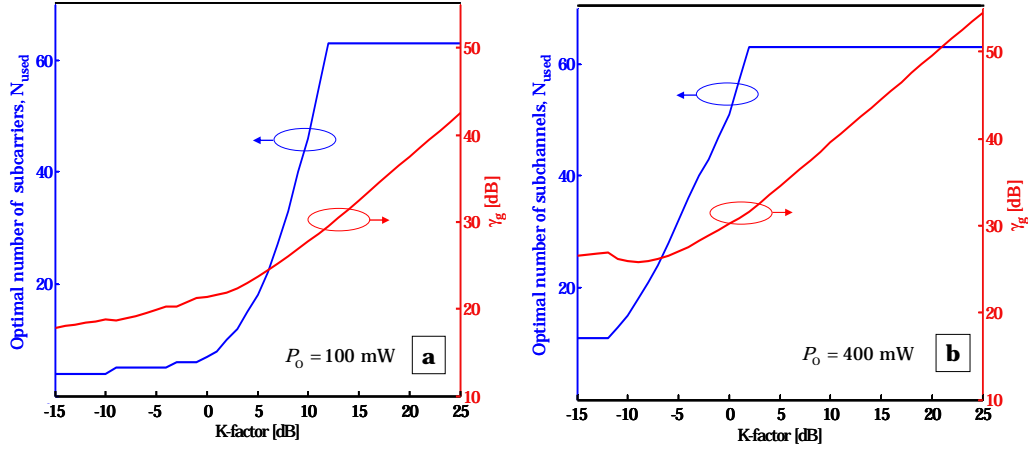


Figure 5.5: The optimal number of subcarriers and the corresponding geometric mean of the parameter γ , for (a) $P_O = 100$ mW and (b) $P_O = 400$ mW.

channel realization), when only a small portion of the bandwidth is used (i.e., the first few subcarriers). This solution is non-adaptive in the way that it is not able to recognize the channel state, so it can not dynamically follow the variations in the channel. This would correspond to the scenario if the channel information is not constantly updated at the transmitter, i.e., if there were no feed-back link (for example, in the system in Fig. 3.8).

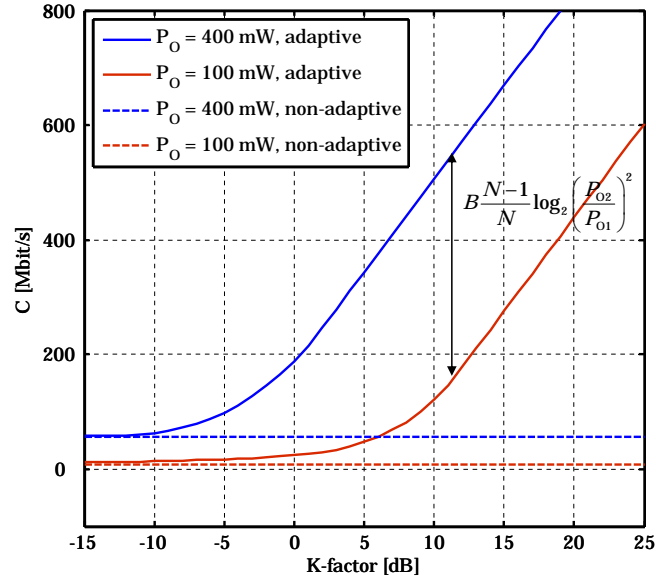


Figure 5.6: Upper capacity bound of an adaptive and a non-adaptive system, for different channel states and mean optical power limit as parameter.

In Fig. 5.6, it becomes clear that a dynamically adaptive system is more than adequate for indoor IR applications. It is able to recognize and exploit the channel knowledge efficiently, thus using the potential for much higher transmission rates than the non-adaptive system. The benefits (with respect to the non-adaptive system) grow with the K -factor, and become significant earlier (lower K -factor) with higher optical power limit. One can observe a relatively slow start in the increase of capacity, turning however into a much faster growth from the point when the link becomes transparent (i.e., all subcarriers are modulated). Then, the number of used subcarriers remains equal to the maximum ($N - 1$), $C \sim \log_2 \prod_{n=1}^{N-1} |H_n|^2$, that is, C grows basically linearly with K [dB] (for large K -factors). The difference between the two plots with different optical power constraints is in this region constant (independent of the channel state) and is given by

$$\Delta C = B \frac{N-1}{N} \log_2 \left(\frac{P_{O2}}{P_{O1}} \right)^2. \quad (5.43)$$

So far, a constant subcarrier bandwidth $B_{SC} = B/N$ was assumed, with $B = 100$ MHz and $N = 64$. However, even when the system is designed with different (maximum) number of independent subcarriers $N - 1$ (which results in different subcarrier bandwidth), it does not have a significant impact on the resulting capacity, especially for large K -factors. The capacity is given by

$$C = \frac{B}{N} \sum_{n=1}^{N_{\text{used}}} \log_2 \left(\frac{1}{N_{\text{used}} 2^{\frac{1}{2N_{\text{used}}}}} \frac{\eta^2 P_O^2 H_n^2 N}{N_0 B} \right). \quad (5.44)$$

As N varies, parameter γ_n varies as well, so it can be expected that the optimal number of the used subcarriers also changes accordingly. Figure 5.7 shows how the optimal number of used subcarriers relates to the maximum number of subcarriers in approximately constant manner. This practically means that about the same portion of the total bandwidth B will be used for transmission, regardless of the maximal number of subcarriers fixed by system design. This is easily verified especially for large K -factors, where $N_{\text{used}} = N - 1 \gg 1$ and the capacity is given by

$$\begin{aligned} C &= B \frac{N-1}{N} \log_2 \left(\frac{N}{(N-1) 2^{\frac{1}{2(N-1)}}} \frac{\eta^2 P_O^2 H_g^2}{N_0 B} \right) \\ &\approx B \log_2 \left(\frac{\eta^2 P_O^2 H_g^2}{N_0 B} \right), \end{aligned} \quad (5.45)$$

which does no longer depend on N .

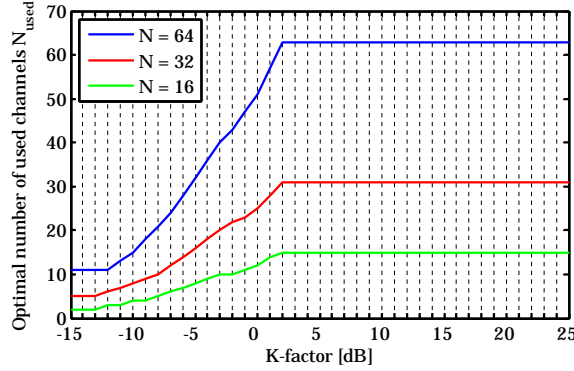


Figure 5.7: Optimal number of used subcarriers for different maximum number of independent subcarriers $N - 1$ within the bandwidth $B = 100$ MHz. The result shows that approximately the same bandwidth is used, no matter how we divide the total bandwidth B , i.e., no matter how large B_{SC} is. As a consequence, according to (5.45) the resulting capacity is practically independent on the granularity of B .

5.2.3 Conservative, straight-forward constraint

The trigonometric moment space approach yields an upper bound of the system capacity for large SNRs. This approach determines the area inside the $2(N - 1)$ -dimensional space $\mathbb{R}^{2(N-1)}$ where the complex sequences (i.e., the OFDM input) need to lie in order to satisfy the condition in (5.22). It is however not obvious how this constraint can be considered in a practical system (with a bit-loading mechanism, for instance), which will be considered in Chapter 6. The intuitive, straight forward condition to ensure

$$1 + \sum_{n=1}^{N-1} (a_n^1 \cos(n\omega_0 t) - b_n^1 \sin(n\omega_0 t)) \geq 0, \quad (5.46)$$

is to fulfill

$$\sum_{n=1}^{N-1} (|a_n^1| + |b_n^1|) \leq 1. \quad (5.47)$$

Upper bounding the system capacity under this simplified constraint is the subject of this subsection, since it will be used as a basis for a capacity curve in a practical system with conservatively designed bit-loading (as shall be seen in Chapter 6). Inequality (5.47) is clearly more restrictive than the true constraint (5.46), thus more conservative capacity results are expected with this approach compared to TMS one.

As for the TMS method, we define a $2(N - 1)$ -dimensional real-valued space inside of which the allowed input elements must lie as

$$\mathcal{SQ}_{N-1} = \{ [X_n^1 = a_n^1 + jb_n^1]_{n=1}^{N-1} : \sum_{n=1}^{N-1} |a_n^1| + |b_n^1| \leq 1 \}. \quad (5.48)$$

Hereby, all inputs are mutually orthogonal, i.e., independent. In order to obtain the upper capacity bound, it is first necessary to determine the volume of \mathcal{SQ}_{N-1} . The bound (5.48) actually defines a $2(N-1)$ -dimensional *cross-polytope* (CPT). A cross-polytope (also known as orthoplex, or hyperoctahedron) is a regular, convex polytope that exists in any number of dimensions, with vertices at all permutations of $(\pm 1, 0, 0, \dots, 0)$. The cross-polytope is the convex hull of its vertices [87]

$$\mathcal{CPT}_k = \left\{ (t_1, \dots, t_k) \in \mathbb{R}^k, \quad \sum_{i=1}^k |t_i| \leq 1, \quad (\forall i) \right\}. \quad (5.49)$$

Figure 5.8 illustrates 2- and 3-cross-polytope.

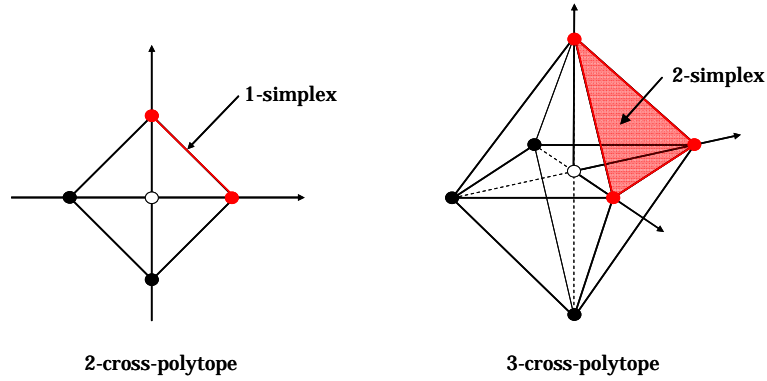


Figure 5.8: Illustration of a (a) two- and (b) three-dimensional cross-polytope, together with the simplices (shaded red) that bound them.

Again, for high SNR, *uniformly* distributed points inside the \mathcal{SQ}_{N-1} provide the capacity upper bound. Therefore, in this case, the entropy of the output signal is given by

$$\begin{aligned} h([\mathbf{Y}]) &\leq \log_2 \left((N\eta P_O)^{2(N-1)} \prod_{n=1}^{N-1} |H_n|^2 \right) + h([\mathbf{X}_{\text{uniform}}^1]), \\ &= \log_2 \left((N\eta P_O)^{2(N-1)} \prod_{n=1}^{N-1} |H_n|^2 \right) + \log_2 V_{\mathcal{SQ}_{N-1}}, \end{aligned} \quad (5.50)$$

where the residual entropy is considered negligible and left out from the expression. Volume of the cross-polytope is denoted by $V_{\mathcal{SQ}_{N-1}} = V_{\mathcal{CPT}_{2(N-1)}}$ and the noise entropy is given by (5.34).

First, the volume of the k -dimensional cross-polytope is determined. It is known that a k -cross-polytope consists of 2^k $(k-1)$ -simplices, [88]. A standard (unit)

k -simplex is defined as [87]

$$\mathcal{S}_k = \left\{ (t_0, t_1, \dots, t_k) \in \mathbb{R}^{k+1}, \quad \sum_{i=0}^k t_i = 1, \quad 0 \leq t_i \leq 1 \quad (\forall i) \right\} \quad (5.51)$$

and presents a k -dimensional analogous of a triangle. For example, a 0-simplex is a point, a 1-simplex is a line segment, a 2-simplex is a triangle, a 3-simplex is a tetrahedron, etc. The vertices of a standard simplex consist of all permutations of $(+1, 0, 0, \dots, 0)$. A standard 1- and 2-simplex are marked in red color in Fig. 5.8. A 2-cross-polytope is bounded by 2^2 line segments, i.e., 1-simplices, and a 3-cross-polytope is bounded with 2^3 2-simplices, i.e., triangles. Therefore, the volume of a k -cross polytope is calculated over the volume *under* a $(k-1)$ -simplex (i.e., between the origin and the simplex in \mathbb{R}^k) [88]

$$V_{\mathcal{CPT}_k} = 2^k V_{\mathcal{S}_{k-1}} = \frac{2^k}{k!}, \quad (5.52)$$

where volume under a standard $(k-1)$ -simplex is known to be $1/k!$. Volume of the space \mathcal{SQ}_{N-1} is then

$$V_{\mathcal{SQ}_{N-1}} = V_{\mathcal{CPT}_{2(N-1)}} = \frac{2^{2(N-1)}}{(2N-2)!} \approx \frac{e^{2(N-1)}}{\sqrt{4\pi(N-1)}(N-1)^{2(N-1)}}, \quad (5.53)$$

when Stirling approximation is applied.

The capacity bound is found as

$$\begin{aligned} C &= B_{\text{SC}} \log_2 \left((\eta N P_O)^{2N_{\text{used}}} \prod_{n=1}^{N_{\text{used}}} |H_n|^2 \right) + \log_2 V_{\mathcal{SQ}_{N_{\text{used}}}} - \frac{1}{2} \log_2 (2\pi e N_{\text{SC}})^{2N_{\text{used}}} \\ &= N_{\text{used}} B_{\text{SC}} \log_2 \left(\frac{\eta^2 P_O^2}{\frac{N_0}{4} B_{\text{SC}}} \frac{e}{8\pi N_{\text{used}}^2} \frac{1}{(4\pi N_{\text{used}})^{1/2 N_{\text{used}}}} \right) + B_{\text{SC}} \log_2 \prod_{n=1}^{N_{\text{used}}} |H_n|^2 \quad (5.54) \\ &= N_{\text{used}} \frac{B}{N} \log_2 \left(\gamma_g \frac{e}{8\pi N_{\text{used}}^2} \frac{1}{(4\pi N_{\text{used}})^{1/2 N_{\text{used}}}} \right), \end{aligned}$$

where N_{used} is determined (like in the previous subsection regarding the trigonometric moment space method) as the the combination of subcarriers which renders the highest capacity. As in the case of (5.38), (5.54) is based on the volume of the space where the input sequences are permitted, but it reveals no information on the particular dimensions (a_n or b_n).

Figure 5.9 shows the capacity obtained by (5.54) for different maximum number of independent subcarriers (i.e., subcarriers carrying independent data, with each

subcarrier containing data in two dimensions) $N - 1 = 15, 31, 63$, while the other parameters are the same as for previous calculations. It can be seen that, as opposed to the results of the trigonometric moment space method (which are provided in the figure for comparison), the capacity increases significantly with the subcarrier bandwidth (smaller N). As $N_{\text{used}} \rightarrow (N - 1)$, (5.54) renders $C \sim -N \log_2 N$. The main reason is that each added dimension is more constrained than the previous ones, but even more than in the TMS case, due to the nature of the constraint (5.38) (all variables are “in phase”, their values add linearly). Moreover, the volume of the cross-polytope decreases with number of dimensions (with respect to the unit-body in the equidimensional space).

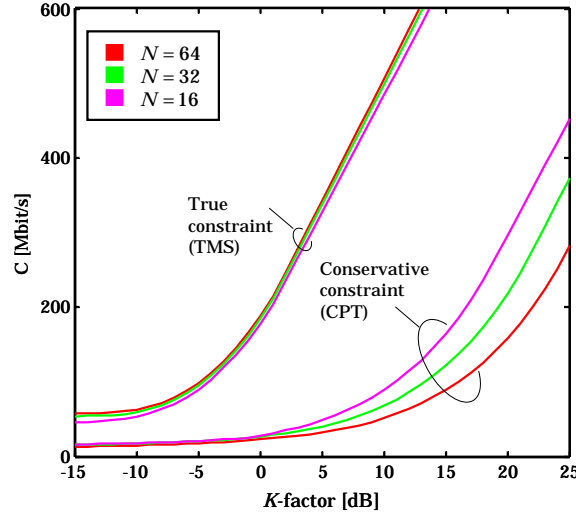


Figure 5.9: Capacity of a system with $N - 1$ independent subcarriers, when complying to the conservative constraint. Assumed parameters: $P_O = 400$ mW and $B = 100$ MHz ($B_{\text{SC}} = B/N$).

Clearly, the resulting capacity plots lie considerably below the upper bound of the trigonometric moment space method, with the most pronounced difference for intermediate K -factors (i.e., where the usable portion of bandwidth grows). Namely, as the number of used subcarriers grows, probability that they all add in phase (which is the worst case approach followed in (5.54)) becomes extremely small, stressing the unnecessary rigorousness of such constraint. Since the conservative condition makes sure that the actual constraint (5.46) is fulfilled, certainly $\mathcal{SQ}_{N-1} \subseteq \mathcal{CH}_{N-1} \subseteq \mathcal{M}_{N-1}$. Therefore, for high SNRs, where the capacity is calculated over the volume inside the space curves, satisfying the conservative constraint leads to a conservative capacity upper bound.⁸ In other words, this conservative bound presents the upper limit for a system which, for the sake of simplicity,

⁸Note, however that, since a uniform distribution was assumed and the noise was neglected for high SNRs, this is not the lower capacity bound.

chooses to satisfy the most conservative constraint. The true system capacity may, however, be larger, as shall be seen from the following example.

5.2.4 Geometric interpretation on a 2D example

Fourier coefficients of non-negative (band-unlimited) functions lie inside a convex trigonometric moment space. This presents a geometric characterization of the condition on the input sequence $[X_n^1]_{n=1}^{N-1}$ (with $X_0^1 = 1$) defined in (5.23). It is convenient to introduce an equivalent algebraic characterization in order to facilitate visualization of the trigonometric moment spaces in case of small number of dimensions. It is known that Fourier coefficients of non-negative functions form a *positive semidefinite sequence* [89]. A sequence $[X_n^1]_{n=0}^{N-1}$ is positive semidefinite if and only if its *Toeplitz matrix*⁹

$$[\mathbf{X}_n^1] = \begin{bmatrix} X_0^1 & X_1^1 & \cdots & X_{N-1}^1 \\ X_{-1}^1 & X_0^1 & \cdots & X_{N-2}^1 \\ \vdots & \vdots & \ddots & \vdots \\ X_{-(N-1)}^1 & X_{-(N-2)}^1 & \cdots & X_0^1 \end{bmatrix} = \begin{bmatrix} X_0^1 & X_1^1 & \cdots & X_{N-1}^1 \\ X_1^{1*} & X_0^1 & \cdots & X_{N-2}^1 \\ \vdots & \vdots & \ddots & \vdots \\ X_{N-1}^{1*} & X_{N-2}^{1*} & \cdots & X_0^1 \end{bmatrix} \quad (5.55)$$

is a positive semidefinite matrix for all $0 \leq n \leq N-1$. Because of the conjugate symmetry ($X_{-n}^1 = X_n^{1*}$), matrix $[\mathbf{X}_n^1]$ is a *Hermitian*¹⁰ as well. For determination of positive semidefiniteness of a Hermitian matrix, it is necessary and sufficient to prove that all the leading principal minors are non-negative [90]. Hence, it is needed that

$$\begin{aligned} \det([\mathbf{X}_0^1]) &= \det[X_0^1] \geq 0 \\ \det([\mathbf{X}_1^1]) &= \det \begin{bmatrix} X_0^1 & X_1^1 \\ X_1^{1*} & X_0^1 \end{bmatrix} \geq 0 \\ \det([\mathbf{X}_2^1]) &= \det \begin{bmatrix} X_0^1 & X_1^1 & X_2^1 \\ X_1^{1*} & X_0^1 & X_1^1 \\ X_2^{1*} & X_1^{1*} & X_0^1 \end{bmatrix} \geq 0 \\ &\vdots \end{aligned} \quad (5.56)$$

for $0 \leq n \leq N-1$. Since $X_0^1 = 1$, the system (5.56) becomes

$$\begin{aligned} X_0^1 &= 1 \geq 0 \\ 1 - X_1^1 X_1^{1*} &= 1 - \|X_1^1\|^2 \geq 0 \\ 1 - 2\|X_1^1\|^2 - \|X_2^1\|^2 + X_1^1 X_1^{1*} X_2^1 + X_1^{1*} X_1^1 X_2^1 &\geq 0 \\ &\vdots \end{aligned} \quad (5.57)$$

⁹Toeplitz matrix of a sequence $[X_n^1]_{n=0}^{N-1} = [X_0, X_1, \dots, X_{N-1}]$ is a square $N \times N$ matrix \mathbf{X} with a property that the element at the position (n, k) is equal to X_{k-n} [90].

¹⁰Hermitian matrix is a square matrix with a property $X(n, k) = X(k, n)^*$ [90]. Toeplitz matrix is a Hermitian if $X_{n-k} = X_{k-n}^*$.

These conditions define the interior regions of the moment spaces.

For the purpose of illustration and clarity, an example of a two-dimensional moment space is considered. Even though the conditions are derived and generally valid for complex-valued (QAM) sequences $[X_n^1]$, real-valued sequence is assumed in this two-dimensional example without loss of generality, i.e., the condition is

$$1 + a_1 \cos(\omega_0 t) + a_2 \cos(2\omega_0 t) \geq 0. \quad (5.58)$$

For the sake of simplicity, it is assumed that $P_O = 1$ and FFT scaling with $2N$ is omitted, so that $a_n^1 = a_n$ and $X_n^1 = a_n/2$. In this case, the conditions (5.58) on X_n^1 render

$$\begin{aligned} -1 &\leq X_1^1 \leq 1 \\ 2(X_1^1)^2 - 1 &\leq X_2^1 \leq 1 \end{aligned} \quad (5.59)$$

so the corresponding trigonometric moment space \mathcal{M}_2 presents a truncated parabola shown in the Fig. 5.10. In the figure, the axis are given for both X_n as well as scaled for a_n . It can be observed that the range for the second dimension depends on the value of the variable in the first dimension. In fact, every added dimension will be more limited than each of the previous ones.

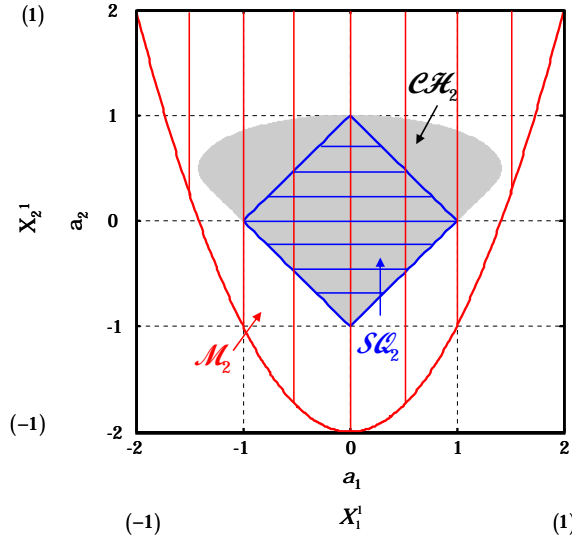


Figure 5.10: Geometrical interpretation and relation of discussed spaces (trigonometric moment space and cross-polytope) on a two-dimensional example.

All possible $[X_n^1]_{n=1}^{N-1}$ sequences (where $N-1 = 2$) which satisfy (5.58) will lie within this moment space. However, the truncated moment space has already been said to present the upper bound of the actual space \mathcal{CH}_2 , meaning that not for all

points within the truncated parabola (5.58) is true. To make this obvious, and to see how tight the calculated upper bound is, a systematic (brute force) check of the condition (5.58) over all the points in space is done. The area where (5.58) is satisfied is shaded gray in Fig. 5.10. It lies inside the truncated parabola, but is considerably smaller, meaning that the bound resulting from the trigonometric moment space method is not tight for low number of dimensions. Nevertheless, as the number of subcarriers $N - 1$ grows, especially in the case of QAM (where the number of dimensions is $2(N - 1)$), the sets \mathcal{CH}_{N-1} and \mathcal{M}_{N-1} approach the same limit, so for $N \rightarrow \infty$ the upper bound becomes asymptotically exact [49].

The third curve shown in Fig. 5.10 presents a bound of a two-dimensional cross-polytope (square), corresponding to the conservative condition, introduced in the previous subsection. In the two-dimensional case, the conservative condition is

$$|a_1| + |a_2| \leq 1 \Leftrightarrow |X_1^1| + |X_2^1| \leq \frac{1}{2}, \quad (5.60)$$

here again with $P_O = 1$ assumed and FFT scaling omitted. When the conservative condition is satisfied, the actual constraint (5.58) is also fulfilled, and we can state that $\mathcal{SQ}_2 \subseteq \mathcal{CH}_2 \subseteq \mathcal{M}_2$. In the two-dimensional case given in Fig. 5.10, the conservative upper bound (volume of the cross-polytope) will be closer to the exact capacity than the actual upper bound (volume of the trigonometric moment space), but it can be expected that with increase of dimensions the actual upper bound becomes the tighter one.

5.3 Performance of water-filling in MSM-based OW systems

As already mentioned, the upper bound derived in Section 5.2.2 is valid for the high SNR region and is a result of rather complex analysis. The water-filling algorithm, introduced in Section 5.1.2 is well known and in power-limited channels leads to true capacity (not only its bound) regardless of SNR. This is why it can be occasionally encountered in the literature as an approximative method for capacity evaluation of an optical IM/DD system (e.g., [91]). In this section, compromises and relaxations of the initial problem, required to properly apply the water-filling algorithm in an optical IM/DD system are explained, and the resultant estimation of system capacity is discussed.

An important fact is that the constraints on the mean optical power and waveform non-negativity (5.14) are bounding the signal in the time-domain. The water-filling approach regards the system capacity over the individual capacities of the parallel subchannels, i.e., it determines the optimal input distributions in the frequency domain. In the case of a power-limited system (constraint on the second moment

of the input distribution), this is not an issue since the Fourier transform is unitary with respect to the second moment (power) of the signal (Parsevals theorem) [56]. This means that the constraint on the total signal power in the time domain is the same as in the frequency domain. However, since the Fourier transform does not share the same property with respect to the first moment of the signal, there is not a unique meaningful statement which can be made on the constraint for the input distributions of the individual subchannels in the IM/DD case.

It would be therefore necessary to translate the existing constraints (5.14) into a constraint on the second moment. This is generally possible when the distribution of the signal in front of the optical source is known. (However, as already mentioned - the overall optimal input distribution is not known). The water-filling algorithm is based on Gaussian input distributions $X_i \sim \mathcal{N}(0, \sigma_n^2)$, which are optimal for the power-limited systems. The signal in front of the optical source is then also Gaussian distributed with the variance equal to the sum of the individual variances

$$x \sim \mathcal{N}(E\{x\}, P_E), \quad P_E = \sum_{n=1}^{N-1} \sigma_n^2. \quad (5.61)$$

Clearly, the mean value constraint can be satisfied by setting $E\{x\} = P_O/S$ (P_O being the limit on the mean optical power, and S the O/E conversion factor at the source). However, since a normal distribution is not a finite one, a strict non-negativity constraint can never be fulfilled, thus the obtained solution is always an approximation. In other words, in case of a Gaussian distribution, some clipping of the signal energy at the optical source is inevitable. The amount of clipping ρ depends on the signal (electrical, modulating) power P_E (variance of x), i.e., on the ratio between the standard deviation ($\sqrt{P_E}$) and the mean value of the distribution (P_O/S)

$$\rho = P(x \leq 0) = 1 - \Phi\left(\frac{P_O}{S\sqrt{P_E}}\right), \quad (5.62)$$

where $\Phi(A) = P(t \leq A) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^A e^{-t^2/2} dt = 0.5 + 0.5\text{erf}(A/\sqrt{2})$ is the cumulative distribution function of a normalized normal distribution. This is illustrated in Fig. 5.11. Since setting a certain allowed clipping probability is equivalent to setting a maximum ratio permitted between the standard deviation and the mean value

$$\frac{\sqrt{P_E}}{(P_O/S)} \leq m_{\text{eff}}, \quad (5.63)$$

the limit on the signal modulating power can be expressed by

$$P_E \leq \frac{P_O^2 m_{\text{eff}}^2}{S^2}, \quad (5.64)$$

where $m_{\text{eff}} \in \mathbb{R}$ can be interpreted as the effective modulation depth, introduced in Chapter 4.1.2.

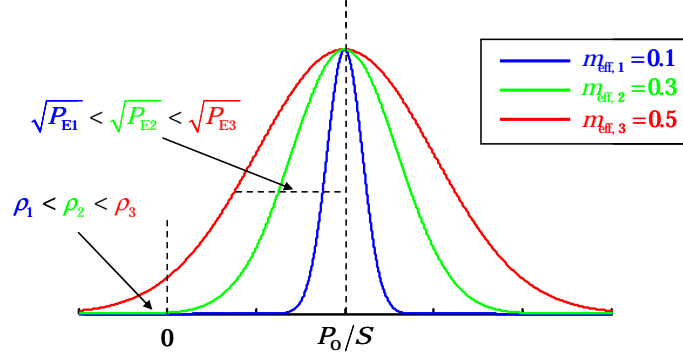


Figure 5.11: Probability of a negative variable (clipping probability) resulting from different ratios between the standard deviation and the mean value of a Gaussian distribution.

From Fig. 5.11 it is clear that there is a trade-off between the accuracy of the approximation and the constraint on the total electrical power. Namely, for a good approximation of the input distribution to Gaussian, the probability of the negative signal should be kept very low. (Otherwise, the input distribution does not resemble Gaussian sufficiently and the capacity obtained using Shannon formula for the subchannel capacity is far overestimated). On the other hand, a relatively low clipping probability reflects adversely on the maximum permissible signal power, as given by (5.64). Hence, even though the needed translation of the first moment constraint into the second moment constraint can be obtained, using an infinite input distribution (such as Gaussian), leads to an unnecessarily conservative power limit. It may actually be much more relaxed if some finite input distribution is considered.¹¹ This is yet another way to see that Gauss is not the optimal (capacity achieving) distribution for the considered channel.

In the following, performance of the water-filling method is discussed. The first challenge is to find an acceptable clipping tolerance, so that the approximation is good enough. As a guideline, one potential tolerance is derived based on the amplitude bound of a practical square-M-QAM-based OFDM system, where with a moderate number of subcarriers, the signal in front of the optical source resembles clipped Gaussian distribution very well. One illustration is given in Fig. 5.12, where $N - 1 = 63$ subcarriers were assumed with $M = 64$ order of QAM. A twofold representation is given - as a probability density function, pdf, (Fig. 5.12 a), and as

¹¹For example, Shannon has shown that the capacity upper bound for peak-power constrained systems (case similar to the system considered here) is achieved with a binary distribution for low SNR, and uniform distribution for high SNR [81].

a normal probability plot¹² of the signal in front of the optical source (Fig. 5.12 b), which allows a better insight in the regions far away from the mean value. A normally distributed data would render a linear plot, and as can be seen from Fig. 5.12 b, an OFDM signal conforms to such distribution, with some discrepancies at the far ends, which come from the finite QAM signal constellations. In the case of smaller number of subcarriers, the discrepancies would increase, but the general conclusions remain the same.

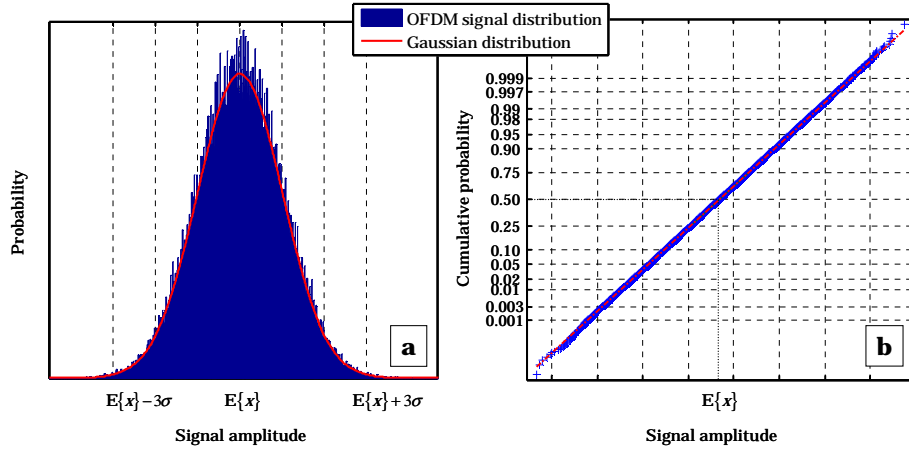


Figure 5.12: Illustration of an OFDM (63 subcarriers with 64-QAM) signal distribution at the optical channel input and its similarity (discrepancy) to the Gaussian distribution. (a) Probability distribution function plot, (b) Normal probability plot.

The constraints of the IM/DD optical channel can be interpreted as a constraint on the maximum signal amplitude. With M-QAM symbol defined as $X = L[(2m_1 - 1) + j(2m_2 - 1)]$ ($= a + jb$ from (3.24)), with $m_1, m_2 \in \left\{-\frac{\sqrt{M}}{2} + 1, \dots, \frac{\sqrt{M}}{2}\right\}$ and a normalizing constant L , the peak amplitude of the OFDM signal (3.24) in front of the optical source is upper bounded by

$$\max |x(t) - E\{x\}| = \frac{2}{2N} \sum_{n=1}^{N-1} \sqrt{2}L_n \left(\sqrt{M_n} - 1 \right), \quad (5.65)$$

where $\sqrt{2}L_n (\sqrt{M_n} - 1)$ is the maximum amplitude in the square constellation of the n^{th} subcarrier (equal to maximum A_n from (3.24)). This case would occur if all the independent subcarriers are in phase, which practically happens with a extremely low probability. However, the only way to ensure no clipping in the

¹²The normal probability plot is a graphical technique for assessing whether or not a data set is approximately normally distributed. The data are plotted against a theoretical normal distribution in such a way that the points should form an approximate straight line. Departures from this straight line indicate departures from normality.

practical system is to require

$$\frac{2}{2N} \sum_{n=1}^{N-1} \sqrt{2} L_n \left(\sqrt{M_n} - 1 \right) \leq \frac{P_O}{S}. \quad (5.66)$$

On the other side, the mean modulating power of the signal in front of the optical source (i.e., again including the scaling of the IFFT) is

$$P_E = \frac{2}{(2N)^2} \sum_{n=1}^{N-1} \frac{2}{3} L_n^2 (M_n - 1). \quad (5.67)$$

For large M_n (≥ 16), $(M_n - 1) = (\sqrt{M_n} - 1)^2 \frac{\sqrt{M_n} + 1}{\sqrt{M_n} - 1} \approx (\sqrt{M_n} - 1)^2$, so

$$\begin{aligned} P_E &\approx \frac{2}{3(2N)^2} \sum_{n=1}^{N-1} 2 L_n^2 \left(\sqrt{M_n} - 1 \right)^2 \\ &\leq \frac{2}{3(2N)^2} \left(\sum_{n=1}^{N-1} \sqrt{2} L_n \left(\sqrt{M_n} - 1 \right) \right)^2 \leq \frac{P_O^2}{6S^2}, \end{aligned} \quad (5.68)$$

where the last inequality results when (5.66) is introduced.

For application of water-filling, one can therefore assume Gaussian distributed signals in all subchannels, which in front of the optical source superimpose to a Gaussian distributed signal $\mathcal{N}(P_O/S, P_O^2/(6S^2))$. From (5.63) and (5.64), this corresponds to $m_{\text{eff}} = 1/\sqrt{6}$ and a clipping probability of $\rho \approx 0.7\%$. However, because of the Hermitian symmetry in the system, only half of the total signal AC power is distributed among the independent channels.

Figure 5.13 shows the capacity results of the water-filling algorithm for 3 different clipping probabilities. This figure is analogous to Fig. 5.6, the result of the trigonometric moment space is repeated for the case $P_O = 400$ mW (and $S = 1$ W/A), here together with several results of water-filling with the same limit on the mean optical power (mind the log-scaled y -axis). For the clipping tolerances, a previously derived value of $\rho \approx 0.7\%$ based on "no clipping" in the practical square M-QAM system is given, as well as one larger ($\rho \approx 10\%$) and one smaller ($\rho \approx 0.1\%$) clipping tolerance (since it is not necessary to assume an M-QAM based system, it is only necessary to have a Gaussian distribution at the channel input).

Comparing the three water-filling results, it is clear that with more clipping permitted (for the same mean optical power constraint), total signal power grows (Gaussian distribution in front of the optical source is wider) and since the Shannon formula does not regard distribution distortion through clipping or any energy loss, the resulting transmission rate grows as well. Nevertheless, at the same time,

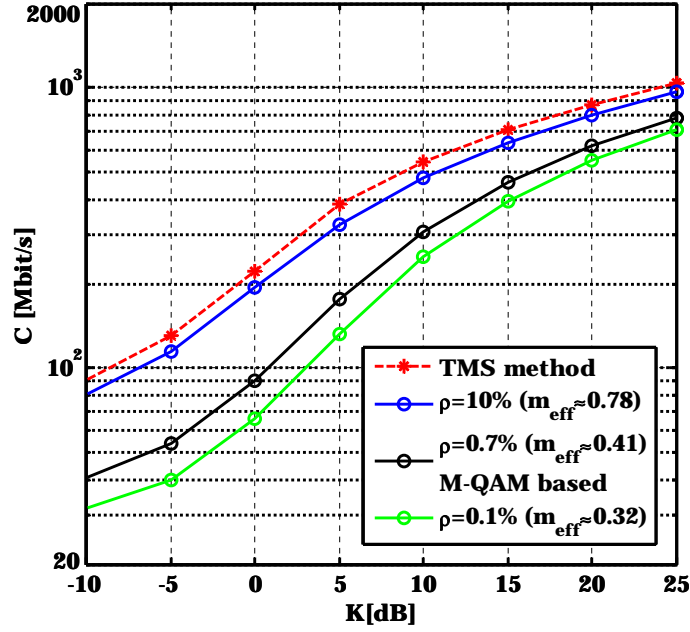


Figure 5.13: System capacity rendered by water-filling (WF) process for three different assumed clipping tolerances, shown together with the trigonometric moment space method upper bound, all for $P_O = 400$ mW. Other assumed parameters: $S = 1$ W/A, $B = 100$ MHz, $N = 64$. Clipping probability is given by (5.62).

the actual distribution which enters the optical channel resembles the Gaussian to a lesser extent, and the approximative result of the water-filling process is increasingly overestimated. In Fig. 5.13, it can also be observed that even the solution of 10% clipping (which basically is of little practical interest¹³), lies below the upper bound obtained by trigonometric moment space. The reason is that the constraint for the water-filling is too conservative as a result of requirement for the infinite Gaussian distribution.

¹³As shall be seen in the Chapter 6.2, clipping probabilities up to a few percent result in error rates correctable by standard mechanisms.

Chapter 6

Throughput Maximization in Practical Systems

The capacity analysis conducted in Chapter 5 is performed in accordance with information theory, under idealized assumptions of error-free transmission and infinitely small information quantum. In practice, however, continuous allocation of resources as well as error-free transmission (use of infinitely long error-correcting codes) are not realizable. In this chapter, the focus of investigations shifts to the throughput maximization in a practical system. Subcarriers of such a system are modulated by square M-QAM and performance is additionally constrained by a permitted error rate. The goal of investigations is to show the potential of efficient signal processing in the considered communication scenario. To that end, Section 6.1 considers throughput maximization by using a loading algorithm which allocates modulation orders and powers to the subcarriers, depending on the current channel state. Furthermore, permitting some clipping of the signal in front of the optical source is investigated in Section 6.2. It is the simplest way to improve the power-efficiency of the MSM system considered, but further enhancements of rate performance are also achievable. Signal processing considered in Chapter 6, as well as a proper data recovery require knowledge of the channel influence on the transmitted signal. Section 6.3 discusses the time interval which is needed for the update of system parameters.

6.1 Rate enhancement using loading algorithms

Achievable throughput in a certain system depends on many parameters. In Chapter 4, both transmitter and receiver components have been characterized, rendering a set of favorable parameters of a realistic system with simple optoelectronic front-ends. Baseband signal bandwidth $B = 100$ MHz is assumed for broadband transmission. This leads to a sampling interval of $T_{\text{sam}} = 5$ ns, so that typical channel lengths discussed in Chapter 2 can be handled with guard intervals, e.g., $L = 16$ samples long (cyclic prefix). In order to have a good spectral efficiency

$(2N/(2N + L))$, $N = 64$ independent subcarriers (including DC) are considered over the bandwidth B . With such values for bandwidth and FFT block, the potentials for throughput maximization with efficient signal processing are investigated in this chapter.

6.1.1 Bit- and power-loading concept

Capacity analysis from Chapter 5 is performed in accordance with information theory. Information theory approach rendered a capacity upper bound for a power- and band-limited system, under idealized assumptions of error-free transmission and infinitely small information quantum. In practice, however, continuous allocation of resources (as by, e.g., water-filling) as well as error-free transmission (the use of infinitely long error correcting codes) are not realizable. Information quantum is given in integer number of bits,¹ depending on the lowest modulation order foreseen. Moreover, performance of each practical solution is limited by the maximum BER allowed in the system (which could be interpreted as the guaranteed quality-of-service in its broadest sense). When a finite set of modulation formats is considered, the achievable system throughput is limited by the highest constellation complexity foreseen.

If the channel can be estimated (at the receiver) and this estimate sent back to the transmitter², the transmission scheme can be adapted relative to the channel characteristics. In practical systems, therefore, the amount of transmitted information (system throughput, which is not the capacity by the information theory definition) can be maximized by deployment of *loading algorithms* (LA) at the transmitter (as previously shown in Fig. 3.8). Based on the channel information, these schemes vary the transmitted power level, constellation size and/or code rate or type of signals transmitted, so that a maximum throughput is achieved, while keeping the system constraints, regarding BER and signal power (or amplitude), satisfied.³ Thus, without increasing the probability of error, these schemes provide high average spectral efficiency by transmitting a lot of information under favorable channel conditions, and reducing the throughput as the channel degrades.

There are several practical constraints which determine when rate-adaptive approach should be used. If the channel is changing faster than it can be estimated,

¹In principle, pulse position modulation (PPM) or similar schemes are able to transmit a part of one bit per channel use. They are valuable for transmission links with very low SNRs, where high transmission rate is not of primary importance.

²The channel coefficients are sent to the transmitter over a feed-back link, which is discussed in Section 6.3.

³Such schemes are often called *rate-adaptive*. The other type of optimization techniques, often encountered in literature are the *margin-adaptive* schemes. There, the goal is to minimize the transmit power for a given rate and BER, or to minimize the BER for a given rate and signal power, as in, e.g., [92].

adaptive techniques may perform poorly, and additional means of mitigating the degradation effects should be used. But, as shall be seen in Section 6.3, in indoor optical wireless channels this is not the case. Finally, hardware constraints may dictate how often the transmitter can change its rate and/or power.

The “time-domain approach” for rate-adaptive links considers adaptive-coding and/or modulation for serial transmission. For OW systems, various proposals, based on adaptive coding with PPM and OOK were proposed in [34–38]. Combining the rate-adaptive transmission with diversity techniques was considered in [39], and with a special self-orienting receiver in [40]. However, baseband modulation schemes use wide-band signals and are easily susceptible to ISI due to multipath propagation. They hence require a complex time-domain equalization at the receiver.

In this work, potentials of adaptive transmission are investigated for a real-valued OFDM-based OW system, where the rate of each individual subcarrier is adapted to the channel conditions (“frequency-domain approach”), with a goal to maximize the total link throughput. Such system deals with channel dynamics and multipath dispersion, both at the same time. The adaptive coding is not considered in this work (only adaptive modulation), but could be a point of further investigation. Adaptive modulation for OW links was recently considered in [93], where the focus was on resource distribution among multiple users, with implementation of a suboptimal loading.

General principles of a loading algorithm are illustrated schematically in Fig. 6.1, for the case of conventional power-limited system. As previously mentioned, the system complies with a predefined maximum BER. On one hand, the choice of the modulation format defines complexity of the symbol constellation and it is well known that a symbol of certain complexity requires a certain (electrical) SNR at the receiver, in order to be detected with a given BER. On the other hand, due to the channel nature, each subchannel n is characterized by a certain equivalent noise level N_n (after reconstructing the transmitted signal), introduced, for example, in (5.13). The noise distorts the transmitted signal and may cause errors in the demodulation process. When the equivalent channel noise increases, the two possibilities to keep the desired BER are either to increase the symbol power or to reduce the symbol complexity. However, in a system with a constrained input, signal power cannot be increased infinitely to compensate for noise. This is also the case in the OW system where the mean optical power and signal sign are limited (which can be interpreted as the amplitude limit, i.e., a peak-power limit). Hence, for each subcarrier (parameterized with the current equivalent noise) individually, a modulation format (i.e., number of bits in one symbol period and therefore data-rate), is chosen from the predefined ensemble, so that the total rate is maximized and yet the power constraint (or its equivalent) is satisfied.

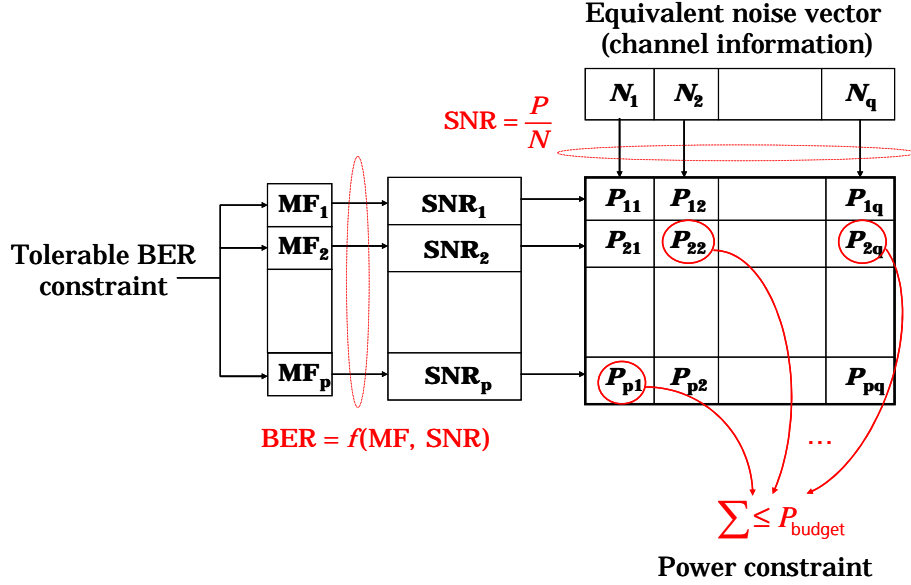


Figure 6.1: Schematic illustration of general loading algorithm principles in multicarrier power-limited systems. Gaussian noise and channel state determine the equivalent noise power N_n on each of q system subcarriers. The choice among p modulation formats (MF) for each subcarrier needs to consider both the BER and the power constraint.

Because of the idealized assumptions, the theoretical analysis provides a benchmark hardly achievable for practical solutions, which can be approached with a certain penalty, depending on the implementation. With this respect, a loading algorithm is considered as optimal or suboptimal. The second criterion for its evaluation is the speed of convergence to the unique solution. There are more than several known bit-loading algorithms developed for application in DSL systems. Some of them provide an optimal solution, but converge slowly (e.g., [94]), and some are computationally efficient but suboptimal (e.g., [95] and [92]). An optimal algorithm proposed by Krongold et al. [96] is chosen to be used in this thesis, since it is able to converge to the optimal solution with convergence time at least as good as the two mentioned suboptimal algorithms.

The Krongold algorithm, in its original form, maximizing the throughput in systems where the second moment of the transmit signal is limited (electrical power), is shortly reviewed in Section 6.1.2. This algorithm is later on deployed to investigate the potentials of adaptive transmission in an OW link. To that end, possible adjustments to strictly comply with the first moment (optical power) and non-negativity constraints on the transmit signal waveform are regarded in Section 6.1.3.

6.1.2 Krongold algorithm for power-limited systems

The Krongold algorithm was developed for application in DSL systems, considering binary phase shift keying (BPSK) and QAM for symbol mapping. It uses the Lagrange multiplier method to maximize the *discrete*⁴ function of the total system throughput (i.e., spectral efficiency) under the total signal power constraint

$$\max(R_{\text{TOT}} = \sum_{n=1}^{N-1} R_n) \quad \sum_{n=1}^{N-1} P_n \leq P_{\text{bud}}, \quad (6.1)$$

where $N - 1$ denotes the number of subcarriers (carrying information) in the system, $R_n = \log_2 M_n$ represents the (integer) number of bits allocated to the n^{th} subcarrier, e.g. via an M-QAM modulation, and P_n represents the constellation power (second moment, electrical power) which is needed for symbol complexity R_n to be detected with a desired BER. The total (electrical) power budget in the system is denoted by P_{bud} .

In fact, depending on the BER required, each MF requires a certain SNR, and depending on the noise in the certain channel this SNR requires a certain power of the transmitted signal, rendering therefore a relation between the rate R_n and the power needed for transmission P_n . Therefore, each variation of bit allocation represents a point in a total-rate-vs-total-power plane ($R_{\text{TOT}}-P_{\text{TOT}}$ plane, where $R_{\text{TOT}} = \sum_n R_n$, $P_{\text{TOT}} = \sum_n P_n$), as shown in blue color in Fig. 6.2 a. This figure presents an illustrative example where allocation over 4 subchannels is assumed with possibilities for 0, 1, 2, 3 and 4 bits.

The Lagrange method with a multiplier $\alpha \in \mathbb{R}$ finds the maximum of the function

$$\begin{aligned} J(\alpha, P_1, \dots, P_{N-1}) &= R_{\text{TOT}}(P_1, \dots, P_{N-1}) \\ &\quad - \alpha \{P_{\text{TOT}}(P_1, \dots, P_{N-1}) - P_{\text{bud}}\} \\ &= \sum_{n=1}^{N-1} R_n(P_n) - \alpha \left(\sum_{n=1}^{N-1} P_n - P_{\text{bud}} \right), \end{aligned} \quad (6.2)$$

by solving the set of equations

$$\frac{\partial J}{\partial P_n} = 0 \quad \Rightarrow \quad \frac{\partial R_n}{\partial P_n} = \alpha \quad (\forall n = 1, \dots, N-1). \quad (6.3)$$

Namely, rates and powers for each subchannel of the optimal solution correspond to the “optimal operating point” on the convex hull of the $R_{\text{TOT}}-P_{\text{TOT}}$ diagram. The optimal operating point is the first point on the convex hull which is met by a line with a slope α . The optimal solution is always on the convex hull, because

⁴Differently to water-filling, where the capacity-vs-total-power function is continuous, due to the infinitely small information quantum.

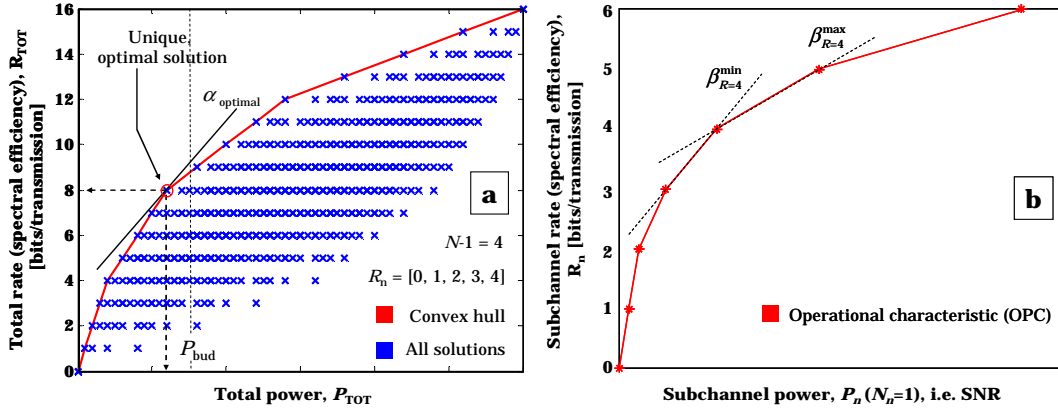


Figure 6.2: (a) R_{TOT} - P_{TOT} diagram with all rate-power combinations, (b) The operational characteristic (OPC), i.e., R_n - $P_n|_{N_n=1}$ curve on one (each) subchannel. Spectral efficiency is denoted by [bits/transmission] which is equivalent to [bits/channel use] or [bits/s/Hz].

this curve connects solutions (bit combinations) requiring a minimum total power for each total rate possible (see Fig. 6.2 a). All other bit combinations lie on the right side of these best solutions, meaning that they require more power. Moreover, because only discrete rates are possible, each point on the convex hull has a continuous range of optimal α values.

On subcarrier n , $SNR_n = P_n/N_n$, so the optimal solution (6.3) can be written over the so-called *operational characteristic* (OPC) of each channel

$$\frac{\partial R_n}{\partial SNR_n} \frac{1}{N_n} = \alpha \quad \rightarrow \quad \frac{\partial R_n}{\partial SNR_n} = \beta_n \quad (\forall n = 1..N), \quad (6.4)$$

where $\beta_n = N_n \alpha$ is the optimal slope on the OPC of subchannel n . The operational characteristic is given as the rate-vs-power characteristic with *unity* equivalent noise (i.e., rate-vs-SNR). This characteristic (illustrated in Fig. 6.2 b) is the same for all the subchannels, and R_n - P_n characteristics have all the same shape, the only difference is the scaling of the OPCs x -axis with N_n .⁵

Conveniently, since OPC is the same for all subcarriers, ranges of optimal β slopes for each operating point on the OPC are constant (as illustrated for $\beta_{R=4}^{min}$ and $\beta_{R=4}^{max}$ in Fig. 6.1.2 b), and can be placed in a look-up-table (LUT). For a particular α , β_n is calculated for each subchannel, and on account of its value, a corresponding

⁵Note that a common α on the R_{TOT} - P_{TOT} diagram translates into an individual β_n on the OPC depending on the noise power in the particular subchannel. This way, instead of handling a unique α in combination with different R_n - P_n characteristics, the problem is transformed into handling a unique OPC (generated independent of noise in a particular subchannel) with different optimal slope parameters β_n .

point (R_n, SNR_n) is read from the LUT. Based on this information, the needed powers are calculated as $P_n = N_n \text{SNR}_n$, and then checked against the total power budget. Finding the optimal α value is proposed to be done iteratively by a *bisection method*, which uses two previously evaluated slopes α_{low} and α_{high} which correspond to $P_{\text{TOT},\text{low}}$ and $P_{\text{TOT},\text{high}}$, such that $P_{\text{TOT},\text{low}} \leq P_{\text{bud}} \leq P_{\text{TOT},\text{high}}$. The bisection method converges to optimal α by calculating the new slope as

$$\alpha_{\text{new}} = \frac{R_{\text{TOT},\text{high}} - R_{\text{TOT},\text{low}}}{P_{\text{TOT},\text{high}} - P_{\text{TOT},\text{low}}}. \quad (6.5)$$

The total power with α_{new} is then evaluated, and the update is made. If $P_{\text{TOT},\text{new}} < P_{\text{TOT}}$ then α_{low} is updated with α_{new} , and if $P_{\text{TOT},\text{new}} > P_{\text{TOT}}$ α_{high} is updated with α_{new} . This procedure is repeated until the new power equals either one of the two starting points.

6.1.3 Conservative approach to loading in an OW system

Since the optical wireless channel is not power-limited in a conventional manner, the Krongold loading algorithm can not be applied in its original form, as in the previous subsection. The reason is that based on the constraints on the signal in front of the optical source (5.14), the electrical modulating power of the signal cannot be uniquely determined.

As already mentioned in Chapter 5.3 (see (5.66)), the only way one can ensure fulfilling the conditions concerning the signal non-negativity and mean value (mean optical power) is to set the maximum amplitude of the signal lower or equal to the DC value (which is, in the best case, set to the optical power constraint). As before, without loss of generality, conversion factor at the optical source is assumed $S = 1$ W/A, and scaling of the FFT blocks is left out. In a system of $N - 1$ independent subcarriers, this condition is equal to

$$2 \sum_{n=1}^{N-1} A_{n,\text{max}} \leq \frac{P_0}{S}, \quad (6.6)$$

where $A_{n,\text{max}}$ presents the maximum amplitude of a signal constellation on the n^{th} subcarrier. Originally, the operational characteristic presents R_n - P_n relationship with unity equivalent noise. In the case where the limit is set on the sum of maximum amplitudes of the constellations on different subcarriers, an operational characteristic R_n - $A_{n,\text{max}}$ with unity N_n is needed.

The relationship between the SNR on a subcarrier modulated by MF_n and its maximum amplitude is given by

$$\text{SNR}_n = f_{\text{max}}(\text{MF}_n) \frac{A_{n,\text{max}}^2}{N_n}, \quad (6.7)$$

where the factor reflecting the ratio between the mean constellation power and the square of its maximum amplitude is

$$f_{\max}(\text{MF}_n) = \begin{cases} 1, & \text{BPSK,} \\ \frac{1}{3} \frac{\sqrt{M_n}+1}{\sqrt{M_n}-1}, & \text{square-M-QAM.} \end{cases} \quad (6.8)$$

In the case of M-QAM constellations where M is an odd integer, there is no common relationship, $f_{\max}(\text{MF})$ depends on the constellation layout. For the purpose of OPC generation, factors for 8-QAM, 32-QAM, and 128-QAM have been found to be 0.599, 0.588 and 0.428, respectively. Because of the quadratic relationship of SNR_n and $A_{n,\max}$, it is convenient to define a normalized SNR for each MF as $\text{SNR}^* = \text{SNR}/f_{\max}(\text{MF})$ and obtain the operational characteristic using the relation

$$\frac{\partial R_n}{\partial A_{n,\max}} = \frac{\partial R_n}{\partial \sqrt{\text{SNR}_n^*}} \frac{1}{\sqrt{N_n}}. \quad (6.9)$$

Therefore, the OPC R_n - $A_{n,\max}$ with a unity effective noise is equivalent to the R_n - $\sqrt{\text{SNR}_n^*}$ relation. Such OPC is presented in Fig. 6.3 a for $R_n = 0, 1, \dots, 8$ and $\text{BER} = 10^{-4}$ as an example.

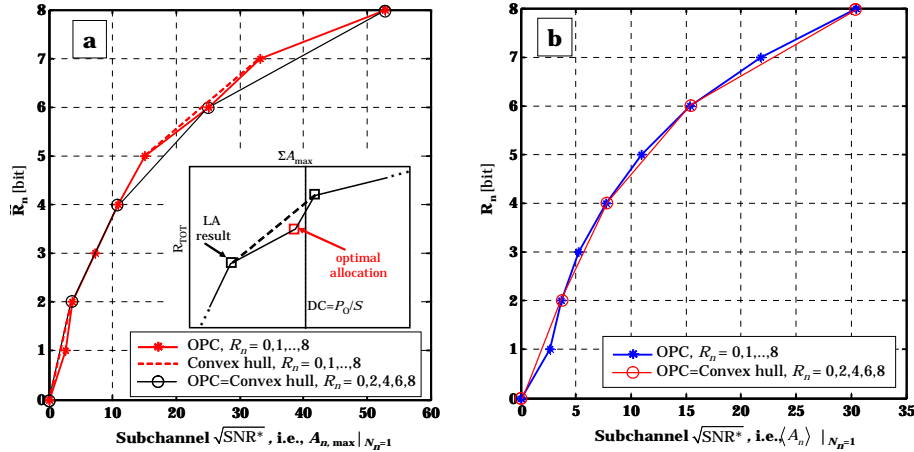


Figure 6.3: (a) The OPC, i.e., R_n - $A_{n,\max}|_{N_n=1}$ plot on one (each) subchannel. (b) The OPC, i.e., R_n - $\langle A_n \rangle|_{N_n=1}$ plot on one (each) subchannel. Changing the BER (in figures 10^{-4}) would not reflect on the shape of the characteristics, only the x -axis would be scaled.

It is clear that the OPC no longer has a concave shape, i.e., that there are some points which lie below the convex hull ($R_n = 1, 6$ are obvious). First consequence of such unfortunate shape is that the algorithm can not be reliably implemented. Namely, it can happen that a slope β_n obtained at some point in the algorithm corresponds to more than one ($R_n, A_{n,\max}$)-pair because the ranges for β (in the LUT)

are not disjunct, which is not foreseen by the implementation, so the algorithm would not converge to a unique solution. Second consequence is that, even if the algorithm converges, a non-concave OPC could lead to a suboptimal solution. In the case of the OPC like in Fig. 6.3 a, the total-rate-vs-sum-of-amplitudes diagram (analogous to the diagram from the Fig. 6.1.2 a) also wouldn't be of a concave shape, and some of the operating points (for some total rates) would lie below the convex hull of that diagram. Since the Lagrange method converges to a point on the convex hull, the LA reliably converges to the optimal solution only if all possible (best) operating points lie on the convex hull of the total $R_{\text{TOT}} - \sum_n A_{n,\text{max}}$ diagram. Since this is not the case, it may happen that the LA “oversees” the optimal solution and converges to the first solution below it on the convex hull, as illustrated in the inset of Fig. 6.3 a. Figure 6.3 b presents the similar plot, with the difference that the SNR is not represented by the maximum constellation amplitude, but with a mean constellation amplitude. The OPC defined in this way will be of interest in the next subsection, however here the same conclusion can be made with respect to the choice of MF set.

The requirement for a concave total-rate-vs-sum-of-amplitudes diagram translates into two conditions for the subchannel OPC: it needs to be concave (that is, to be equivalent to the convex hull of all possible solutions) and to have uniform steps on the y -axis. The term “uniformity of the y -axis” means that the difference between two adjacent considered symbol complexities is constant. The necessity for this second condition can be easily shown by a counterexample (assuming $R_n = 0, 2, 3, 4$ also results in a “non-concave” total diagram). Based on the previous discussion, for implementation of the Krongold algorithm in the OW channel with the limit on the sum of constellation amplitudes, only square QAM are taken in consideration ($R_n = 0, 2, 4, 6, \dots$), which assures reliable convergence of the algorithm to the unique optimal solution under the given conditions.

Figure 6.4 presents the rates obtained by the loading algorithm (LA), in the case when bits are distributed over $N - 1 = 63$ subcarriers and under the limit on the sum of maximum constellation amplitudes (6.6). The total rate R_b [bits/s] is obtained by

$$R_b = \frac{B}{N} R_{\text{TOT}} = \frac{B}{N} \sum_{n=1}^{N-1} R_n, \quad (6.10)$$

which represents the gross rate (without the influence of cyclic prefix or other overhead components). In the figure, three curves are given, reflecting different target BER values. Clearly, the achievable rates can be increased by relaxing the error tolerance. With deployment of, e.g., sophisticated error correction techniques or automatic repeat request algorithms, BER can be set as high as 10^{-3} - 10^{-2} , however, their complexity will influence the needed system overhead. A capacity

upper bound for the conservative LA solution is also presented as a benchmark, with derivation given in the following subsection.

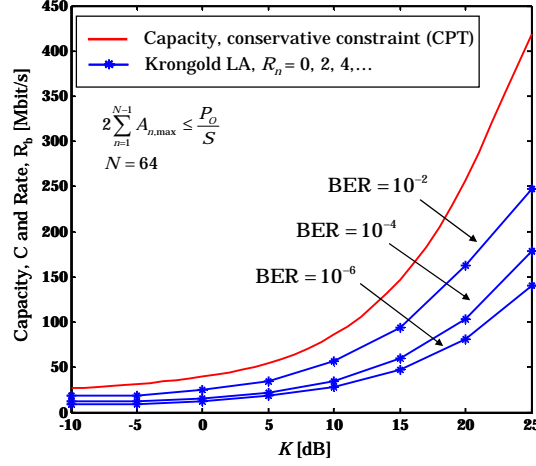


Figure 6.4: Comparison of *conservative Krongold LA* (regarding the sum of amplitudes, as discussed in this subsection) with *conservative capacity bound* (based on the method from Chapter 5.2.3, but with a slightly adjusted constraint, as discussed in Chapter 6.1.4). Assumed parameters: $P_0 = 400$ mW, $S = 1$ W/A and $B = 100$ MHz.

6.1.4 Upper-bounding the conservative loading solution

For the superimposed signal in front of the optical source ((3.24) scaled with $(2N)$, and notation $A_0 = X_0$)

$$X_0 + 2 \sum_{n=1}^{N-1} \sqrt{(a_n^2 + b_n^2)} \cos(2\pi f_n t + \arctan(b_n/a_n)) \geq 0 \quad (6.11)$$

needs to be satisfied. The algorithm implemented in Section 6.1.3 deals with the limit on the sum of maximum constellation amplitudes ($\sqrt{a_n^2 + b_n^2}$), and complies to the constraint

$$\sum_{n=1}^{N-1} \sqrt{a_n^2 + b_n^2} \leq \frac{X_0}{2}, \quad (6.12)$$

which is the straight-forward (intuitive) approach in this case. Since it presumes all the signals on the independent subcarriers to be in phase, it considers a more conservative constraint than (6.11). Derivation of the capacity upper bound can however not be done directly under the condition (6.12), but we can transform this condition so that it can easily be solved by the conservative constraint case introduced in the previous chapter (5.2.3). As shall be seen, the capacity bound

obtained in this way is also an upper bound on (6.12).

Consider introducing a parameter r_n for each n so that $\sqrt{a_n^2 + b_n^2} \leq r_n$ and rewriting the final sum in (6.12) as

$$\sum_{n=1}^{N-1} r_n = \frac{X_0}{2}. \quad (6.13)$$

Each point defined by the pair (a_n, b_n) which lies inside the circle of radius r_n , also lies inside the square $|a_n| + |b_n| \leq r_n\sqrt{2}$ around this circle. This implies that all the solutions a_1, \dots, b_{N-1} of the problem (6.12) lie inside of an $\mathbb{R}^{2(N-1)}$ -space defined by

$$|a_1| + |b_1| + \dots + |a_{N-1}| + |b_{N-1}| \leq \sqrt{2} \sum_{n=1}^{N-1} r_n = \sqrt{2} \frac{X_0}{2}. \quad (6.14)$$

The volume of the cross-polytope defined by (6.14) (with length $\sqrt{2}$ between each vertex and the origin) is therefore larger than of the space inside the space curve (6.12) and can serve to upper-bound the rates obtained by the conservative Krongold LA introduced in the previous subsection. Similarly as in Chapter 5.2.3 (section regarding the conservative constraint), volume of the $2N_{\text{used}}$ dimensional cross-polytope

$$V_{S_{Q_{N_{\text{used}}}}} = \frac{2^{2N_{\text{used}}}}{(2N_{\text{used}})!} (\sqrt{2})^{2N_{\text{used}}}, \quad (6.15)$$

and corresponding capacity upper bound when N_{used} independent subcarriers are used

$$C_{\text{Kron}} = N_{\text{used}} B_{\text{SC}} \log_2 \left(\gamma_g \frac{e}{4\pi N_{\text{used}}^2} \frac{1}{(4\pi N_{\text{used}})^{1/2N_{\text{used}}}} \right). \quad (6.16)$$

Note that in this case, (6.15) presents a volume of a cross-polytope with length $\sqrt{2}$ between each vertex and the origin (whereas in Chapter 5.2.3, a cross-polytope with unit length between each vertex and the origin was considered), which results in an additional factor 2 in the logarithm function argument in (6.16) with respect to (5.54).

This upper limit is given in the Fig. 6.4, as a reference for the rates obtainable by the conservative Krongold LA. It can be observed that, as expected, no matter how powerful correction codes are assumed, the rate found under the condition (6.6) will not achieve this bound. Finding the bound directly under the constraint (6.12) would provide an even tighter upper bound, i.e., the resulting curve would lie somewhat under the curve presented in Fig. 6.4. On the other hand, when the

values are compared to the capacity upper bound derived by the trigonometric space moment method in Chapter 5.2.2 (Fig. 5.6 or even Fig. 5.9 in 5.2.3), the bound (6.16) is significantly lower (and the limit under (6.12) surely even more).

Considering all the signals in phase (practically a non-existent scenario) clearly leads to transmission rates which are well below the actual limits. Nevertheless, there is no other way to upfront ensure the non-negativity of the transmitted waveform based on an M-QAM as the primary modulation alphabet used in the OFDM link. Since practical systems do not consider error-free transmission anyhow, the non-negativity condition can be somewhat relaxed, and additional errors due to clipping of high peaks (only negative or both positive and negative) allowed up to a certain extent. This is in more detail investigated in the following section.

6.2 Controlled clipping

Basic investigations of OW multiple subcarrier (MS) systems, conducted in [41–43], point out poor power efficiency as their major disadvantage. Transmission signal of MS-based systems is namely characterized with relatively seldom but high peaks, which may be translated in requirements for a large DC signal in order to stably drive the optical source. On the other hand, the safety regulations set a limit on the average optical transmission power, i.e., on the DC signal. These conditions can restrict the achievable rates to rather conservative values. Approaches for improvement of power-efficiency (in terms of DC reduction) proposed in literature were based on block-coding [97–99] or peak-to-average-power-ratio reduction techniques originally proposed for RF OFDM systems [100, 101]. Nevertheless, such techniques usually reduce transmission rate and may have large implementation complexity, which limits the number of subcarriers to only a few. Also a variation of DC component on symbol interval basis was proposed, which is not suitable for the adaptive modulation systems (convergence of the loading algorithm is questionable).

An alternative approach, often implemented in practical systems, is to permit some controllable amount of clipping at the transmitter to enhance the power efficiency. A special OFDM scheme which foresees clipping was suggested in [102]. That approach, however, exploits only a half of the channel (baseband) bandwidth, since it assumes only modulation of odd subcarriers. In this thesis, it is assumed that the additional clipping errors could be post-compensated or even tolerated, since the very high peaks occur rather infrequently in the OFDM signal. In a system where DC is limited, if somewhat greater tolerance of error performance is accepted by allowing for clipping, significant improvements in transmission rate can be expected. Equivalently, for the same targeted transmission rate, a significantly lower DC is needed.

The objective of this section is to investigate the influence of controlled clipping on the performance of an OFDM-based OW system (with parameters defined in previous sections) in terms of both error and transmission rate. Hereby, *symmetric* and *asymmetric* clipping of the digital time samples (before digital-to-analog conversion) are both considered. In the former approach, both positive and negative peaks are clipped, whereas the latter regards clipping of the negative peaks only. Differently to the high-frequency amplifiers used in radio systems, dynamic laser characteristic is much more linear in the high power range than the static one, so it could be argued that the negative peaks are more critical in an optical system, whereas the positive peaks could go through with little distortion. Given that fewer samples undergo clipping in the asymmetric approach, less errors are expected. Nevertheless, since the investigation of non-linearities in the Section 4.1 considered carrier-to-interference suppression for a symmetrical signal driving range around DC, the results of the asymmetric clipping approach should be taken as optimistic.

As previously mentioned, if clipping is permitted, the modulating power of the transmission signal can be increased, which leads to higher transmission rates (of course, assuming that the error performance still remains under the tolerance boundary). There are two ways to approach this. The total AC power can be either directly increased, or it can be increased indirectly, by increasing the sum of subcarrier amplitudes until the error performance reaches a certain limit. Since the investigation from the previous section regarding the limit on sum of amplitudes and rate when no clipping is permitted can serve as a basis for further improvement, the same approach is taken in the following investigations to qualitatively evaluate and describe the effects of clipping onto BER. For rate maximization, however, both approaches are considered and compared.

Simulations scenario

The effect of clipping on system performance is investigated by means of Monte-Carlo (MC) simulations. For a certain assumed channel state (described by the K -factor within the range -10 to +25 dB), and a certain signal DC limit assumed (eye safety limit of 400 mW optical power), the sum of constellation amplitudes on $N - 1$ independent subcarriers (denoted as the *amplitude budget*, A_{bud}) is increased in steps over the DC and the resulting total BER is observed. The reference to Fig. 5.13 and relation of BER to clipping probability ρ will be made in Section 6.2.3. For each new A_{bud} , a new result of the loading algorithm is obtained. It is assumed that each subcarrier can carry $R_n = \{0, 2, 4, \dots\}$ bits. The loading algorithm itself distributes bits and amplitudes yielding to a chosen BER, assuming only the presence of noise in the channel (it does not account for clipping errors). This error rate is further on denoted as BER_{LA} , as opposed to the total error rate,

marked by BER_{TOT} , which includes errors due to clipping.

At this point, it is convenient to make two digressions. First, it was necessary to define the amplitude budget in Section 6.1.3 as the sum of *maximum* amplitudes of the constellations associated to the subcarriers (recall (6.6)), in order to absolutely ensure no clipping. Nevertheless, in the following, the amplitude budget is redefined as the sum of the *mean* amplitudes of the constellations⁶ associated to the subcarriers⁷

$$A_{\text{bud}} = \sum_{n=1}^{N-1} \langle A_n \rangle \quad (6.17)$$

and the corresponding operational characteristic (OPC) is given in Fig. 6.3 b. The amount of clipping and resultant error rate depend hence on the ratio of the sum of amplitudes on positive frequencies and signal DC

$$\frac{2 \sum_{n=1}^{N-1} \langle A_n \rangle}{(P_O/S)} = \frac{2A_{\text{bud}}}{X_0}. \quad (6.18)$$

We can assume that when the ratio (6.18) is 1, practically no clipping occurs (even though (6.17) denotes the sum of mean and not maximum constellation amplitudes). The reason is that the statistic of the clipping influence in that case is too low to cause a significant error increase (i.e., observable in simulation results), yet, the rates provided are less conservative. This will be observable in the examples illustrated later on. As the ratio (6.18) increases, more clipping occurs at the source and more errors are expected.

With definition of the amplitude budget as in (6.17), the relationship between the SNR on a subcarrier modulated by a modulation format MF_n and its mean amplitude ($\langle A_n \rangle = \langle \sqrt{a_n^2 + b_n^2} \rangle$, when $X_n = a_n + jb_n$) is given by

$$\text{SNR}_n = \frac{P_n}{N_n} = f_{\diamond}(\text{MF}_n) \frac{\langle A_n \rangle^2}{N_n}, \quad (6.19)$$

where P_n is the mean power of the signal constellation ($P_n = \langle |X_n|^2 \rangle = \langle A_n^2 \rangle$). Factor $f_{\diamond}(\text{MF}_n)$ is the ratio between the mean power and the square of the mean amplitude of the “basic” square M-QAM constellation derived as

$$f_{\diamond}(\text{MF}_n) = \frac{\frac{2}{3}(M_n - 1)}{\left(\frac{4}{M_n} \sum_{p=1,3,\dots}^{\sqrt{M_n}-1} \sum_{q=1,3,\dots}^{\sqrt{M_n}-1} \sqrt{p^2 + q^2} \right)^2}. \quad (6.20)$$

⁶As before, with a complex symbol on the n^{th} subcarrier defined as $X_n = a_n + jb_n$ its amplitude is given by $A_n = \sqrt{a_n^2 + b_n^2}$ and $\langle A_n \rangle$ denotes the mean amplitude of the constellation on the n^{th} subcarrier.

⁷As before, for the sake of simplicity, scaling of FFT operation blocks is omitted. Moreover, the budget is chosen to be defined over positive frequencies.

If the loading algorithm, distributing the mean constellation amplitudes $\langle A \rangle$ and modulation formats M , renders $(\langle A_n \rangle, M_n)$ on each particular subcarrier n , the corresponding total signal modulating power (including the power on the negative frequencies, i.e., the subcarriers carrying conjugated symbols) is given by

$$2P_{\text{eq}} = 2 \sum_{n=1}^{N-1} P_n = 2 \sum_{n=1}^{N-1} f_{\text{D}}(\text{MF}_n) \langle A_n \rangle^2, \quad (6.21)$$

which is in general not equal to $((2A_{\text{bud}})^2)$ since P_{eq} depends on the distribution of $\langle A_n \rangle$ over the subcarriers, and not on the chosen A_{bud} alone. By recalling (5.62) from the Chapter 5.3, clipping probability ρ can be determined as

$$\rho = s \left(1 - \Phi \left(\frac{X_0}{\sqrt{2P_{\text{eq}}}} \right) \right), \quad s = \begin{cases} 1, & \text{asymmetric clipping,} \\ 2, & \text{symmetric clipping,} \end{cases} \quad (6.22)$$

where the factor s (at the beginning of the expression) differentiates the cases of symmetric and asymmetric clipping. Similarly, by recalling (5.63), we can denote the ratio $\sqrt{2P_{\text{eq}}}/X_0$ as the effective modulation depth m_{eff} .

The second digression regards the fact that MC simulations inherently provide a symbol error rate (SER) on individual subcarriers, whereas the parameter of interest here is the bit error rate. If Gray coding [53] is assumed for M-QAM constellations, BER can be calculated as $\text{SER}/\log_2 M$, however only if a symbol is mistaken for one of its closest neighbors (where the error is one bit). Nevertheless, for an A_{bud} high enough, the error vector magnitudes (EVMS) become sufficient to result in a decision which is more than 1 bit “off”, as illustrated in Fig. 6.5 a. With certainty, we can write

$$\text{SER}/\log_2 M \leq \text{BER} \leq 1 - (1 - \text{SER})^{1/\log_2 M}, \quad (6.23)$$

where the right-hand side of the inequality presents the case when all bits within the symbol are falsely interpreted (worst case). By comparison of the two cases for different number of subcarriers (N) and different channel states (K), it is however found that significant discrepancies exist only when BER is in the area higher than 10^{-2} , as shown by the example given in Fig. 6.5 b. Incidentally, the K -factors shown are all positive, but the situation is no different for the negative values. The x -axis shows the ratio between the sum of amplitudes on all subcarriers (including the conjugated ones, hence the factor 2) and signal DC, i.e., X_0 . Up to 0 dB, practically no clipping is expected. On the y -axis, the mean BER over all subcarriers is shown. The maximum BER allowed in the regarded system is assumed in the order of magnitude 10^{-3} , where forward-error-correction (FEC) can be made. Therefore, the estimation of BER from SER using the left-hand expression from (6.23) is concluded justified for the following investigations.

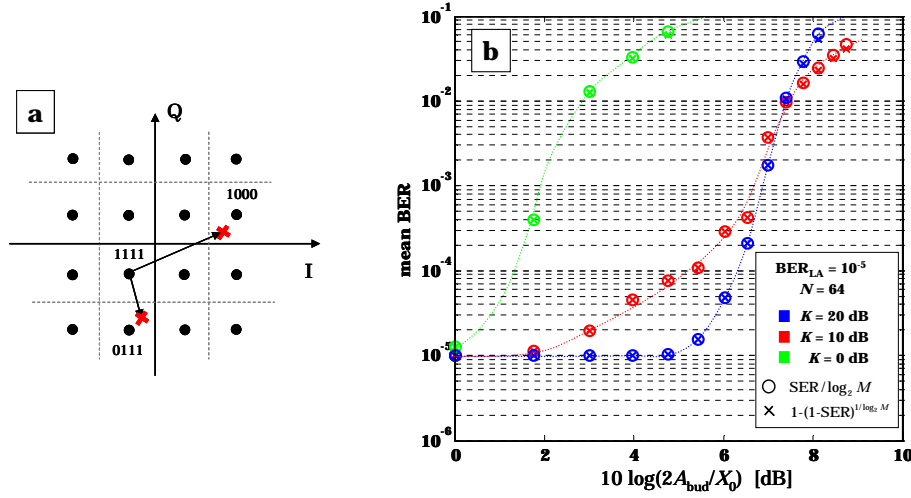


Figure 6.5: (a) Gray coded square 16-QAM constellation when a symbol error has one and two falsely detected bits, (b) Comparison of the minimum and maximum BER resulting from the same SER, for three illustrative K -factors. Assumed parameters are $N = 64$, $P_O = 400$ mW, $S = 1$ W/A. Modulation orders of individual subcarriers result from the loading algorithm (can be different from point to point), but are left out of discussion for now.

6.2.1 Influence on error performance

This subsection investigates how the increase of the signal amplitude budget influences the overall system error performance. The achievable rates under a certain error performance are investigated and discussed in Section 6.2.2, and are in this subsection left out of discussions. In an OFDM-based system with clipping in the digital domain, the signal going into the DAC and to the optical source can be represented as the sum of the signal before clipping x and a clipped signal portion c (also a comb of Kronecker delta functions)

$$x_C(k) = x(k) + c(k) \quad k \in [0, 2N - 1]. \quad (6.24)$$

Hereby, the coefficient $c(k) = 0$ in the case when clipping does not occur, i.e. if $0 \leq x(k) \leq 2X_0$. Figure 6.6 illustrates where in the system the assumed clipping takes place and how the relevant signals look like. For the sake of simplicity, it is considered that the cyclic prefix is inserted after clipping.

As introduced in Appendix B (when discussing the OFDM transmission chain in Chapter 3.3), DFT and IDFT operations can be represented by left multiplication with Vandermonde matrices $[\mathbf{W}_N]$ and $[\mathbf{W}_N]^{-1}$, respectively (B.11). The OFDM symbol in time domain before clipping is in vector-matrix representation therefore given by

$$[\mathbf{x}] = [\mathbf{W}_{2N}]^{-1} [\mathbf{X}], \quad (6.25)$$

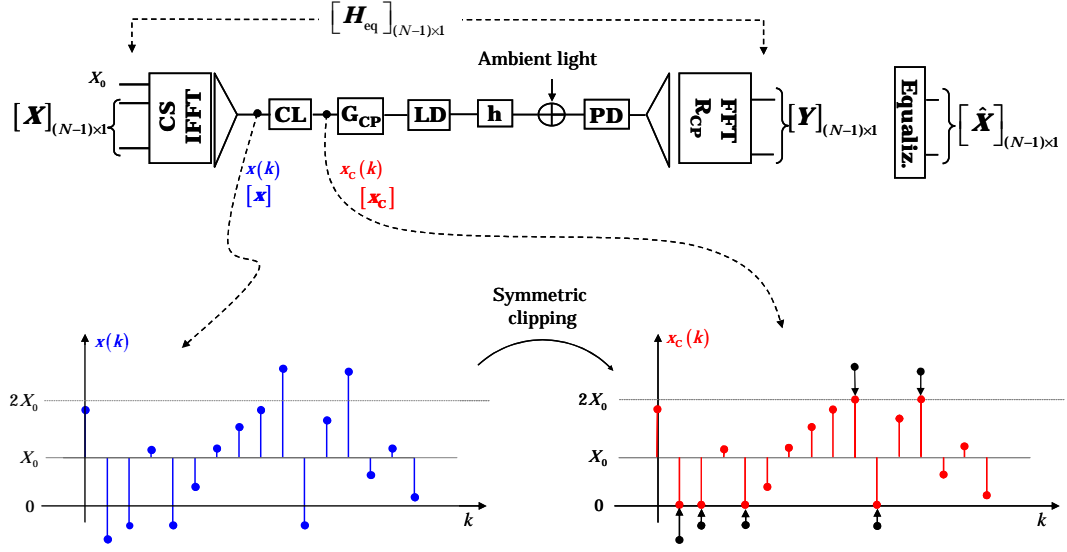


Figure 6.6: Graphical representation of symmetric clipping enforced onto the digital time samples at the OFDM transmitter.

where $[\mathbf{X}]$ denotes the input signal vector in frequency domain (with $2N$ elements). After clipping, $[\mathbf{x}_c] = [\mathbf{x}] + [\mathbf{c}]$. If operations of cyclic prefix insertion and removal are denoted by $[\mathbf{G}_{CP}]$ and $[\mathbf{R}_{CP}]$, respectively, the channel matrix by $[\mathbf{h}]$, and the background noise vector by $[\mathbf{z}]$, the received signal at the system output (in frequency domain) is

$$\begin{aligned}
 [\mathbf{Y}] &= [\mathbf{W}_{2N}] [\mathbf{R}_{CP}] \left([\mathbf{h}] [\mathbf{x}_c] + [\mathbf{z}] \right) \\
 &= [\mathbf{W}_{2N}] [\mathbf{R}_{CP}] [\mathbf{h}] [\mathbf{G}_{CP}] [\mathbf{W}_{2N}]^{-1} [\mathbf{X}] \\
 &\quad + [\mathbf{W}_{2N}] [\mathbf{R}_{CP}] [\mathbf{h}] [\mathbf{G}_{CP}] [\mathbf{c}] + [\mathbf{W}_{2N}] [\mathbf{R}_{CP}] [\mathbf{z}] \\
 &= [\mathbf{H}_{eq}] [\mathbf{X}] + [\mathbf{H}_{eq}] [\mathbf{W}_{2N}] [\mathbf{c}] + [\mathbf{W}_{2N}] [\mathbf{R}_{CP}] [\mathbf{z}].
 \end{aligned} \tag{6.26}$$

The second summand of the last equality denotes the Fourier transform of the clipped portion of the signal x , i.e., the Fourier transform of c , and it is responsible for the influence of clipping on signal degradation. The third summand represents the spectrum of the additive noise. The output at the subcarrier n is therefore

$$\begin{aligned}
 Y_n &= H_{eq,n} \left(X_n + \sum_{k=0}^{2N-1} c_k e^{-j2\pi \frac{k}{2N} n} \right) + Z_n \\
 &= H_{eq,n} \left(X_n + (a_{n,c}^2 + j b_{n,c}^2) \right) + Z_n.
 \end{aligned} \tag{6.27}$$

From (6.27), it can be concluded that each clipped sample causes distortion of the whole signal spectrum. Clipping introduces a random shift of the sent symbol ($a_{n,c}^2 + j b_{n,c}^2$), depending on amplitudes and phases of the symbols sent on all

subcarriers.⁸ It therefore presents an additional interference which can significantly affect the decision process. From the signal-to-noise-and-clipping-interference-ratio (SNCR)

$$\text{SNCR}_n = \frac{\langle |X_n|^2 \rangle}{\frac{\langle |Z_n|^2 \rangle}{|H_{\text{eq},n}|^2} + \langle a_{n,c}^2 + b_{n,c}^2 \rangle}, \quad (6.28)$$

can be seen that unless the variance of the clipping distortion ($\langle a_{n,c}^2 + b_{n,c}^2 \rangle$) is negligible with respect to the background noise, there occurs a rise of BER_{TOT} due to clipping errors.

Figure 6.7 a presents the influence of clipping errors on the BER performance, illustrated for the channel state $K = 15$ dB. For the assumed limit on the mean optical power P_O (due to safety regulations), which is fulfilled by setting $X_0 = P_O/S$, clipping of high peaks is enforced on the digital signal $x(k)$ as

$$x_c(k) = \begin{cases} 0, & x_c(k) < 0 \quad (\text{in asymmetric and symmetric case}), \\ 2X_0, & x_c(k) > 2X_0 \quad (\text{in symmetric case}), \\ x(k), & \text{otherwise,} \end{cases} \quad (6.29)$$

like shown in Fig. 6.6. Bit-error rates in Fig. 6.7 a are given as the mean values (calculated over all active subcarriers), with a bar representing the maximum and minimum BER which occurred across the subcarriers. Since BER_{LA} of 10^{-5} is assumed, in case when clipping is not enforced (imaginary scenario of a perfectly linear infinitely long laser characteristic), the result of the MC simulations reflects the system performance only under disturbance of background noise, corresponding to the error rates targeted by the loading algorithm, no matter how high the chosen amplitude budget is. This (blue) plot therefore serves as a reference, but it is also an indicator that simulations are properly performed.⁹

From the two plots reflecting cases of clipping it can be concluded that the modulating power (amplitude budget) can be increased up to some point without noticeable influence on the total BER, which remains in the range of 10^{-5} set for the loading algorithm. At the same time, with a larger amplitude budget, the amount of information transmitted is growing. In this range, in (6.28), variance $\langle a_{n,c}^2 + b_{n,c}^2 \rangle$ is still negligible with respect to the background noise component. Unavoidably,

⁸Amplitudes and timings of clipping samples $c(k)$ are determined by the data sent and therefore random. The clipped signal component has a multipath character and the more clippings occur, the more the distribution resembles a complex Gaussian one.

⁹In practical systems, it is often implemented that the part of the budget, left over after the distribution, is equally divided among the active subcarriers, in order to enhance the SNR and minimize errors (below BER_{LA}). To be able to distinguish the clipping from noise effects more easily, this step is not included in the simulation results presented in this work. Nevertheless, additional simulations with such distribution of extra budget were performed, which rendered practically the same outcomes, leading to the conclusion that the budget is usually well exhausted.

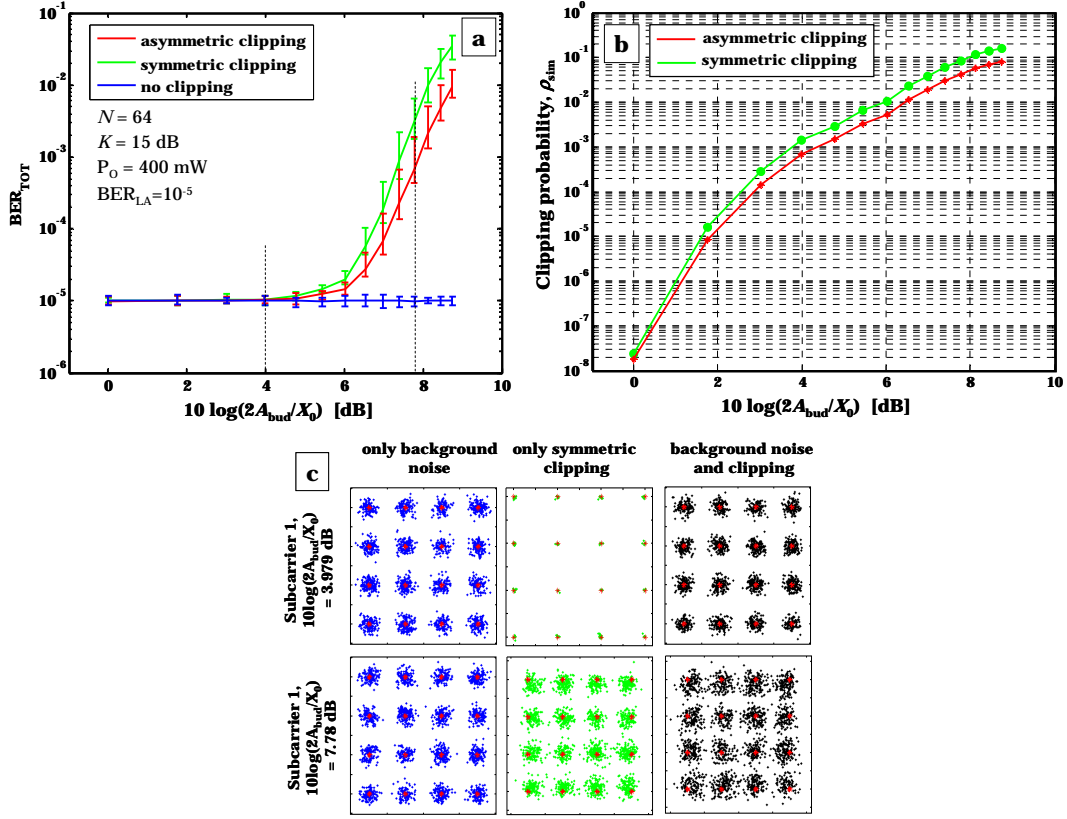


Figure 6.7: (a) Influence of clipping on the rise of errors in the system for an illustrative channel realization. Assumed parameters are $N = 64$, $P_O = 400$ mW and $S = 1$ W/A. Modulation orders of individual subcarriers result from the loading algorithm, but are left out of discussion for now. (b) Corresponding clipping probability. (c) Received signal constellations on the first subcarrier, when errors in detection occur only due to background noise, only due to clipping or both.

as the signal power is further increased, the errors due to clipping become larger and occur more often, evoking a steep increase of BER_{TOT} , which asymptotically approaches 0.5. It can also be observed in the figure that the error rates are always higher in case of symmetric clipping than with asymmetric clipping for the same amplitude budget (since approximately twice as many samples are clipped).

Corresponding clipping probabilities shown in Fig. 6.7 b are calculated as

$$\rho_{\text{sim}} = \frac{\text{Number of clipped samples}}{\text{Total number of transmitted samples}} \quad (6.30)$$

for both symmetric and asymmetric clipping. Even though differently obtained, as a system parameter, ρ_{sim} corresponds to ρ from (6.22) and from Chapter 5.3. As expected, in the symmetric case, clipping probability is always twice as large. It can be observed that the resultant clipping probabilities are quite moderate, and

that they remain below 10% for the BER values of interest (below 10^{-2}). For the following investigations, we consider BER_{TOT} as the relevant design parameter, and assume that the corresponding clipping probabilities lie in the range of several percent. As already mentioned, clipping probabilities are in more detail discussed in the Section 6.2.3.

An illustration of the couple of received signal constellations is shown in Fig. 6.7 c. In particular, a constellation on the first subcarrier (incidently modulated with 16-QAM, other subcarriers can be differently modulated, depending on the result of the loading algorithm) is shown for two amplitude budgets (indicated by dashed lines in Fig. 6.7 a), and reflecting the errors due to background noise alone (in blue, left-hand side), due to symmetric clipping alone (in green, middle) and due to both effects simultaneously (in black, right-hand side). In the case of relatively small amplitude budget (upper panels) the error performance is clearly determined by the background noise, while clipping, although obviously present, has no significant influence. On the other hand, from the bottom panels can be observed that clipping has strong impact on the total error performance when large A_{bud} is assumed. Hereby, the constellations reflecting influence of background noise remain in both cases unchanged, since the loading algorithm in both cases targets the same $\text{BER}_{\text{LA}} = 10^{-5}$.

In cases when a lot of clipping occurs, the error clouds due to clipping resemble the clouds due to background noise, leading to the assumption that clipping can also be modeled as additive Gaussian noise. Figure 6.8 shows the normal plot diagrams for the real and imaginary part of the error signal obtained on the first subcarrier due to clipping errors exclusively, corresponding to the constellation shown in the mid-lower panel of the Fig. 6.7 c. From this figure it can be concluded that errors practically follow Gaussian distribution in both dimensions ($a_{n,c}$ and $b_{n,c}$ are Gaussian distributed), i.e., the error vector from (6.28) has a Rayleigh distributed amplitude and uniformly distributed phase.

This means that if SNCR is known, BER_{TOT} can be calculated by the theoretical expressions known from literature for AWGN channels. This was confirmed by calculating the variance of clipping errors and SNCR for each subcarrier by MC simulations, and comparing to the error rate result of the MC simulations. Incorporating, however, the influence of the clipping effect into the loading algorithm is difficult (if not impossible), since clipping depends on data sent on all subcarriers, unlike the noise of background light which can be processed individually, subchannel by subchannel.

Figure 6.9 presents the BER plots for different channel realizations, which clearly show that the increase of BER with the increase of signal power depends on the channel state. Plots corresponding to both symmetric and asymmetric clipping are

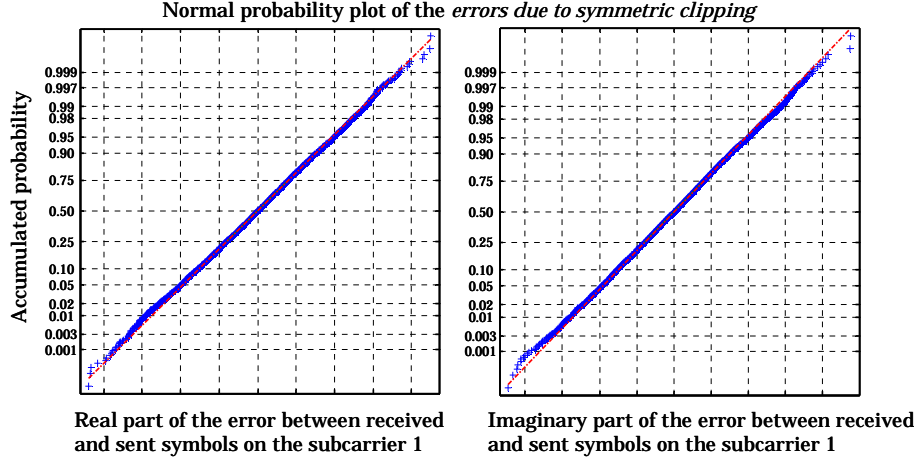


Figure 6.8: Normal probability plot of the errors due to symmetric clipping on the first subcarrier. Mathematical expectation is in the middle of the x -axis.

given. For such behavior, following intuitive explanation can be given. In principle, increasing the signal power results in two effects:

- (a) two bits are allotted to the so far unused subcarriers (all subcarriers “start” with 4-QAM, where all symbols have the same amplitude and power), which has a minor influence on BER or/and,
- (b) amplitudes of already active subcarriers are increased, by increasing the modulation order, which has a strong influence on the resultant BER growth.

When the effect (a) is dominant over the effect (b), the increase of BER is slow and steady, whereas, in the opposite situation, BER increases rapidly. It can be observed in the figure, that in the case $K = 0$ dB, effect (b) is dominant from relatively low A_{bud} , whereas for relatively flat channels there is a significant range of A_{bud} for which the effect (a) is dominant, before (b) takes over. It can be concluded that with a larger K -factor, as long as predominantly new bits are allocated, the increase of BER is slower, at the same time, however, the effect (b) becomes dominant for lower A_{bud} (compare cases $K = 15$ dB and $K = 20$ dB in Fig. 6.9). This can be explained by the fact that the same increase of A_{bud} causes more subcarriers to be activated when K is higher. It is, however, important to realize that for a certain BER tolerance in the system (which permits clipping), maximum amplitude budget (indirectly power budget) allowed is different for different channel realizations. Acceptable BERs in a system depend on the error correction mechanism.

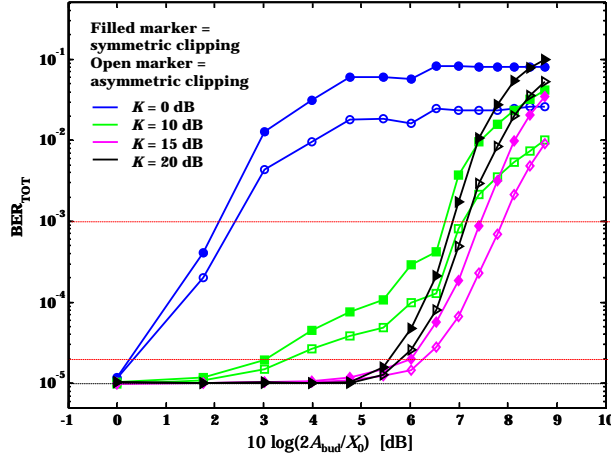


Figure 6.9: Influence of clipping on system error performance for different channel realizations. Assumed parameters are $N = 64$, $P_O = 400$ mW and $S = 1$ W/A. Modulation orders of individual subcarriers result from the loading algorithm, but are left out of discussion for now.

6.2.2 Rate maximization

Further investigations are now conducted to show how some clipping tolerance can improve the system performance in terms of achievable rates. Hereby, it is assumed that loading algorithm aims at $\text{BER}_{\text{LA}} = 10^{-5}$, and that the total BER in the system does not exceed $\text{BER}_{\text{TOT}} = 2 \cdot 10^{-5}$ when clipping errors are included. For each particular channel state of interest, the maximum amplitude budget is found¹⁰, which keeps $\text{BER}_{\text{TOT}} \leq 2 \cdot 10^{-5}$. In the simulations, signal DC value of $X_0 = 400$ mA (i.e., $P_O = 400$ mW with $S = 1$ W/A) is assumed, and the amplitude budget A_{bud} is increased in steps of 25 mA (starting from $A_{\text{bud}} = X_0/2$). This can lead to some inaccuracy in the resultant maximum budget,¹¹ but with minor (insignificant) implications on the achieved rate.

Figure 6.10 a shows the A_{bud} limit over different channel states. Four plots are provided. Two stair-like plots reflect the cases when symmetric (red) and asymmetric (blue) clipping is foreseen in the system. The two horizontal lines represent the cases when clipping is not permitted in the system, and the loading algorithm itself aims at $\text{BER}_{\text{LA}} = 2 \cdot 10^{-5} = \text{BER}_{\text{TOT}}$. No clipping in the system can be achieved either by deploying the *conservative* loading algorithm which distributes the *maximum* constellation amplitudes, with a constraint set as $A_{\text{bud}} = \sum A_{n,\text{max}} = X_0/2$ (black line), or by deploying the loading algorithm which distributes the mean constellation amplitudes (as in all cases considered in this section), but then with such

¹⁰Actually, if the two budgets render the same transmission rate, and both keep the BER below the limit, the lower budget is chosen. This is often the case by relatively poor channels.

¹¹The maximal achievable rates can be slightly higher.

a budget, which ensures no clipping for any channel state of interest (for worst case channel realization). This amplitude budget is then lower than $X_0/2$ and is the same for both “no symmetric clipping” and “no asymmetric clipping” case (green line in Fig. 6.10 a).

From Fig. 6.10 a, it can be concluded that when clipping is permitted, the maximum A_{bud} first increases with channel quality, however decreases again for very large K -factors. The reason is that for such particularly good channels, only a moderate or small budget increase is enough to lead the system into the state where all subcarriers are active. Hence, further increase of A_{bud} results only in increase of modulation orders (effect (b) from previous discussion), causing a rapid BER deterioration. For poorer channel realizations some budget increase over DC is possible, however not a large one, since, in such low-pass-like channels, the extra power conveys in the increase of the amplitudes of the few active subcarriers. Nevertheless, in majority of channel states, for the small decrease of error performance (from 10^{-5} to $2 \cdot 10^{-5}$), system can operate with much more modulating power, and expect to provide significantly increased transmission rates.

This can be observed in Fig. 6.10 b, where the comparison of achievable rates in terms of achievable total spectral efficiencies ($R_{\text{TOT}} = \sum R_n$ [bits/OFDM symbol] as in Chapter 6.1) is given for the same cases considered. Moreover, it can be observed that the application of the conservative algorithm (black plot) results in better rates than deployment of the algorithm distributing the mean amplitudes when no clipping is allowed (green plot).

Figure 6.10 c illustrates the distributions of bits (denoted by different colors of grid-squares) over the subcarriers (x -axis) for different channel states (y -axis) in the case of no clipping with conservative algorithm (R_n distribution corresponds to the black R_{TOT} plot in Fig. 6.10 b). Figure 6.10 d illustrates the distribution in case of symmetric clipping (R_n distribution corresponds to the red R_{TOT} plot in Fig. 6.10 b). From these distributions, it can be observed that with clipping permitted, for high K -factors, 16-QAM and even 64-QAM are often usable, whereas in the case of no clipping, the active subcarriers are (for the observed K -range) modulated with 4-QAM.

So far, it has been assumed that noise and clipping have more or less the same contribution to the total error rate, i.e.,

$$\text{BER}_{\text{TOT}} = \text{BER}_{\text{LA}} + \text{BER}_{\text{CL}}, \quad \text{BER}_{\text{LA}} \approx \text{BER}_{\text{CL}}. \quad (6.31)$$

In the following, different ratios between noise and clipping error rates are investigated, in order to determine the influence on the improvement of transmission rate. Hereby, instead of bit-error-rates, symbol-error-rates are targeted in simulations, because of the slightly better accuracy (recall the digression related to Gray coding

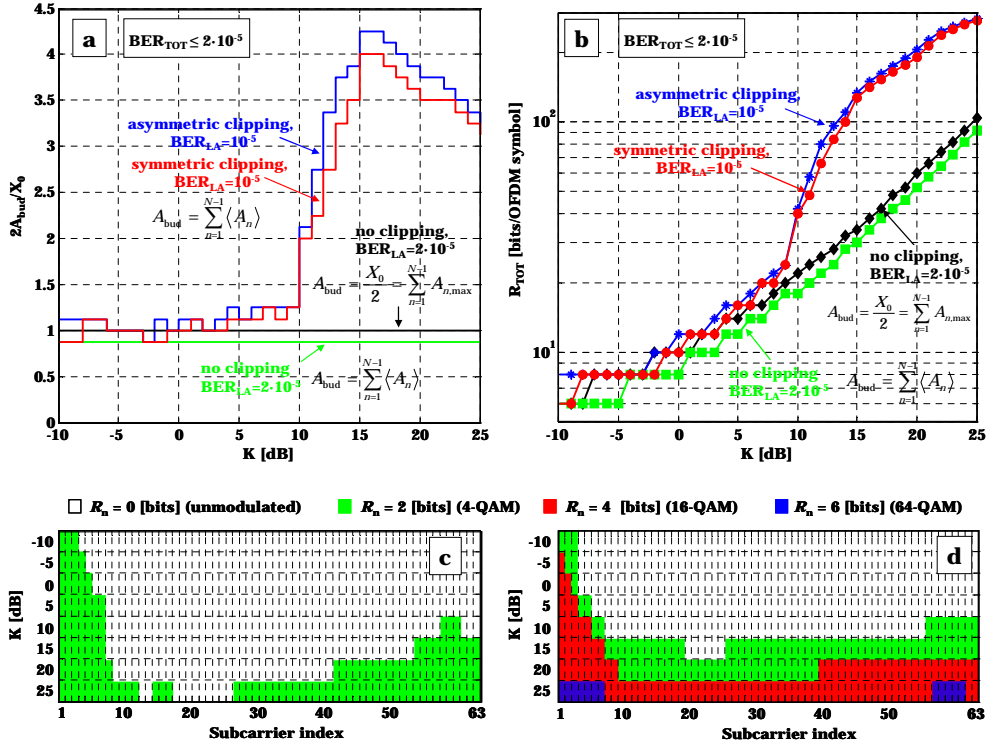


Figure 6.10: Comparison of the (a) permissible sum of constellation amplitudes (with respect to signal DC) and (b) achievable transmission rates, when clipping is implemented and when no clipping is permitted in the system, all for the same overall error performance limit. Figures (c) and (d) show distributions of bits over the subcarriers for various K -factors, in the case of no permitted clipping and distribution of maximum amplitudes (corresponding to the black plot in figure (b)), and permitted symmetric clipping and distribution of mean amplitudes (corresponding to the red plot in figure (b)). Assumed are $N = 64$, $P_O = 400$ mW and $S = 1$ W/A.

at the beginning of this section).¹² The results are presented in Fig. 6.11. In order to have a sound basis for MC simulations, in this case, the targeted SER_{TOT} is set to $2 \cdot 10^{-3}$. Then, instead of BER_{LA} at the input of the loading algorithm, symbol error rate SER_{LA} is targeted.¹³

Figure 6.11 shows the achievable rates (Fig. 6.11 a and Fig. 6.11 c) and permissible amplitude budgets (Fig. 6.11 b and Fig. 6.11 d) for symmetric (two upper panels) and asymmetric clipping (two lower panels) in the cases when

1. $\text{SER}_{\text{LA}} = 1.9 \cdot 10^{-3}$, so that allowed $\text{SER}_{\text{CL}} \leq 10^{-4}$; the ratio between the

¹²Since the conclusions have a rather qualitative character, they are just as valid for practical systems, where BER is targeted.

¹³This leads to different values at the x -axis of the operational characteristic than in Fig. 6.3 b, otherwise the algorithm remains the same.

errors due to noise and errors due to clipping is well in favor of noise (green).

2. $\text{SER}_{\text{LA}} = 10^{-3}$, so that allowed $\text{SER}_{\text{CL}} \leq 10^{-3}$; noise and clipping contributions are approximately the same (red).
3. $\text{SER}_{\text{LA}} = 10^{-4}$, so that $\text{SER}_{\text{CL}} \leq 1.9 \cdot 10^{-3}$; the ratio between the errors due to noise and errors due to clipping is well in favor of clipping (blue).

From the figures representing the achievable rates R_{TOT} , it can be concluded that a significant portion of clipping errors is beneficial. Rates achieved in case $\text{SER}_{\text{LA}} = 1.9 \cdot 10^{-3}$ are clearly below the other two cases, since there is practically no buffer for clipping, and the permissible A_{bud} values are closer to the $X_0/2$, as can be seen from the right-hand side panels.

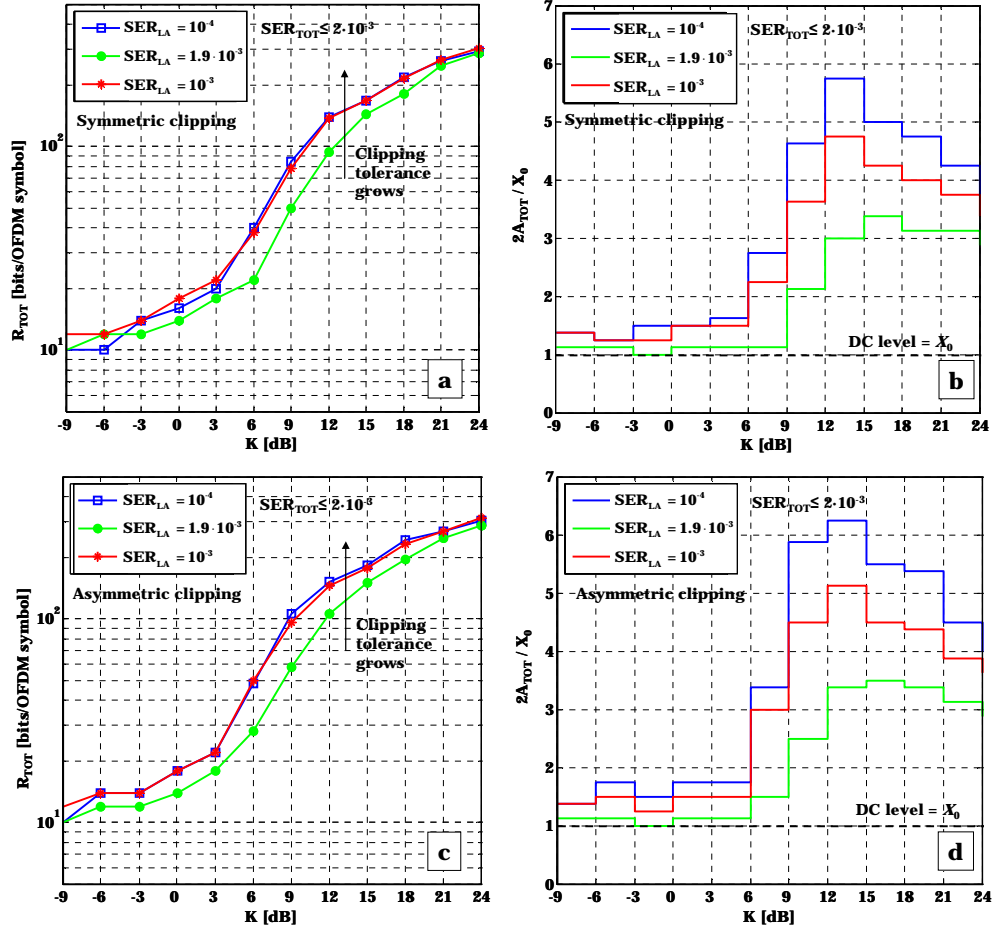


Figure 6.11: Comparison of the (a)(c) achievable transmission rates and (b)(d) permissible sum of subcarrier amplitudes for different contributions of background noise and clipping effect to overall system error performance. Assumed are $N = 64$, $P_O = 400$ mW and $S = 1$ W/A.

Cases 2 and 3 show very similar performances in terms of achievable rates. Setting, however, the SER_{LA} target too conservatively (case 3) requires deployment of higher signal powers for the same performance (see panels b and d). Additionally, the transmission rate R_{TOT} , achieved for a given channel state, increases when A_{bud} increases and/or when SER_{LA} increases ($R_{\text{TOT}} = f(A_{\text{bud}}, \text{SER}_{\text{LA}})$). So, if the LA target is too conservative (case 3), it can happen that the increase in amplitude budget (with respect to DC), gained by higher clipping tolerance, is not enough to cope with the decrease of the allowed noise contribution. Then, no significant improvement of rate can be achieved, or even a decrease with respect to maximum rate occurs. Therefore, it can be concluded that it makes most sense to target the total SER (or BER) with approximately equal contributions due to noise and clipping (case 2).

If the loading algorithm considers a finite set of modulation formats, then once all channels are filled with maximum number of bits, an increase of the K -factor leads to no further increase in rate. But since less power is needed for the transmission of these bits, the error performance gets better and approaches the performance without clipping. The loading algorithm applied in this thesis considered a maximum of 12 bits per subcarrier, and the K -factors for which this occurs lie beyond the range shown in the figure (and beyond the range of interest).

As announced at the beginning of this section, the effect of clipping on error and rate performance in a considered system is so far investigated and qualitatively described by varying the sum of subcarrier amplitudes since it is easier to compare the results to the “no clipping” case. Now, once clipping is permitted in the system and if the total BER limit is the only constraint, in a similar way, one can directly increase the electrical power budget and distribute the electrical power over the subcarriers using the conventional Krongold loading algorithm (from Section 6.1.2).

In order to avoid confusion, at this point it is convenient to recapitulate some relevant guidelines. Irrespective of the loading approach, the optical power constraint P_0 at the source sets a limit on the mean value of the transmitted waveform (e.g., (3.23) or (3.24)) so that $X_0 = P_0/S$. In order to avoid any clipping, the modulating signal waveform should be kept within $[0, 2X_0]$ boundaries. This condition is met by the approach presented in Section 6.1.3, where the loading algorithm was adjusted to distribute the maximum amplitudes of the signal constellations on individual subcarriers ($A_{n,\text{max}}$). In this section (Section 6.2), clipping is permitted at the optical source, which means that the transmit signal waveform occasionally reaches outside of the $[0, 2X_0]$ -range. The two approaches mentioned earlier relate to the parameter chosen to indicate how much outside this range the signal waveform reaches.

The first approach, adopted so far in this section, defines a certain sum of mean

constellation amplitudes on the independent subcarriers A_{bud} (6.17). This amplitude budget is distributed over the independent subcarriers by the loading algorithm from Section 6.1.3, but with the operational characteristic from Fig. 6.3 b, which regards the mean constellation amplitudes. This approach is referred to in further text and in figures as “amplitude distribution” or “amplitude distributing algorithm”. Resultant signal modulating power $2P_{\text{eq}}$ can be obtained by (6.21), the corresponding effective modulation depth by $m_{\text{eff}} = \sqrt{2P_{\text{eq}}}/X_0$ and clipping probability ρ by (6.22).

The second approach assumes a certain total modulating power on the independent subcarriers, defined as the power budget

$$P_{\text{bud}} = \sum_{n=1}^{N-1} P_n. \quad (6.32)$$

Total waveform modulating power is then $2P_{\text{bud}}$ (including the power on negative frequencies). Budget P_{bud} is distributed over the independent subcarriers by the loading algorithm from Section 6.1.2, i.e., by the Krongold algorithm in its original form which results in the optimal P_n distribution. This approach is referred to in further text and in figures as “power distribution” or “power distributing algorithm”. The corresponding effective modulation depth is obtained by $m_{\text{eff}} = \sqrt{2P_{\text{bud}}}/X_0$ and clipping probability ρ by (6.22) where P_{bud} is used instead of P_{eq} .

Figure 6.12 a presents the achievable rates in terms of spectral efficiencies (R_{TOT} in bits/OFDM symbol) when such power distribution is considered (blue plots) and when the so far assumed amplitude distribution is considered (red plots), assuming symmetric clipping. Transmission rate can then be obtained with R_b [bit/s] = $(B/N)R_{\text{TOT}}$, as introduced by (6.10). The results for $\text{BER}_{\text{TOT}} \leq 2 \cdot 10^{-5}$ and $\text{BER}_{\text{TOT}} \leq 2 \cdot 10^{-3}$ are shown, with corresponding loading algorithms aiming at $\text{BER}_{\text{LA}} = 10^{-5}$ and $\text{BER}_{\text{LA}} = 10^{-3}$, respectively. The increase of power budget P_{bud} is performed in simulations in 0.625 mW steps.

In Fig. 6.12 a, it can be observed that for a significant range of channel realizations, distribution of powers results in better rates. Most prominent differences appear for the intermediate K values (not too low and not too high). This occurs because for a particular K from this range, the permissible power budget P_{bud} (6.32) which still satisfies the limit on BER_{TOT} , is larger than the modulating power on independent subcarriers P_{eq} obtained by (6.21) from the amplitude distribution (with the A_{bud} permissible for this K).

This can be seen in Fig. 6.12 b, where total modulating powers of the two approaches are compared. The dashed red plots present the signal power (normalized by X_0^2) when the amplitude distribution loading algorithm is used, i.e., P_{eq} .

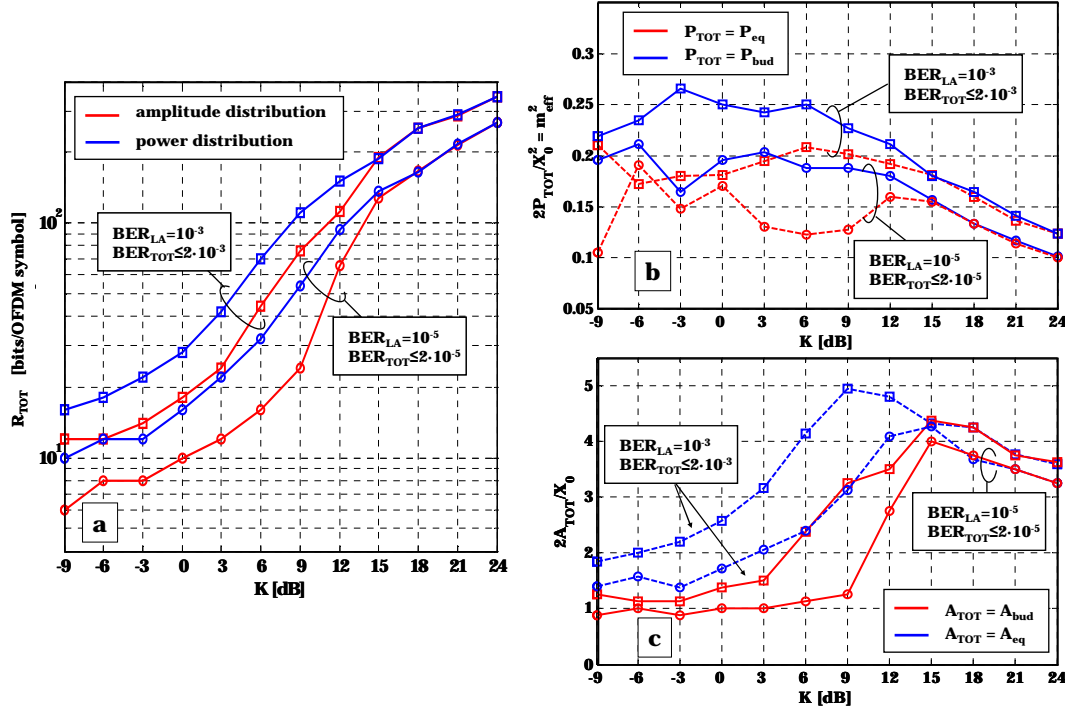


Figure 6.12: Comparison of the (a) achievable rates in terms of bits per OFDM symbol, (b) total signal modulating powers and (c) sums of amplitudes when the loading algorithm with amplitude distribution (red plots) and the loading algorithm with power distribution (blue plots) are implemented and symmetric clipping is enforced. Assumed are $N = 64$, $P_O = 400$ mW and $S = 1$ W/A.

The permissible P_{bud} , resulting from the condition on BER_{TOT} when the power distributing LA is used, is shown (also normalized by X_0^2) by the blue solid plots. Note that the variable P_{TOT} denoted on the y -axis is only auxiliary (see the legend). The results on the y -axis can be interpreted as m_{eff}^2 . It can be seen that for these channels $P_{bud} > P_{eq}$. And this result can be explained in the following way.

In the considered system with permitted clipping, beside the maximum DC value, the limit on BER_{TOT} presents the *only* constraint. (This constraint is not of physical nature, but results from the system design). Since the limit on BER indirectly poses the limit on the signal power (rather than amplitude), it can be expected that the power distributing loading algorithm maximizes the rate. The amplitude distributing algorithm achieves lower rates (see Fig. 6.12 a), because it yields to an *additional* constraint on the sum of amplitudes.

On the other side, similarly as the sum of powers P_{eq} can be obtained from the amplitude distributing approach (6.21), we can obtain the sum of mean amplitudes A_{eq} from the P_n distribution of the power distributing algorithm. If the distribution

a certain P_{bud} results in powers and modulation formats (P_n, M_n) on individual subcarriers, the equivalent sum of amplitudes is

$$A_{\text{eq}} = \sum_{n=1}^{N-1} \sqrt{\frac{P_n}{f_{\text{eq}}(M_n)}}. \quad (6.33)$$

Figure 6.12 c shows a comparison of the maximum permissible sum of amplitudes (including the negative frequencies and normalized by X_0) permitted under a given BER_{TOT} constraint (red solid plots) and the equivalent sum of amplitudes (normalized in the same way), resulting from the distributions of the maximum permitted P_{bud} , obtained from Fig. 6.12 b under the same channel and BER conditions (blue dashed plots). It can be observed that A_{eq} well exceeds the maximum A_{bud} , permitted as the sum of amplitudes under same BER condition. This means that the advantage in rate of the power distributing algorithm is obtained due to disregarding the condition on the sum of amplitudes. Nevertheless, having such additional constraint may be necessary if the crest-factor or amount of clipping are of interest, which is especially important when asymmetric clipping is considered in the system.

According to the simulations result, the difference in rate for K -factors up to about 15 dB, is due to the fact that power distributing LA results in more active subcarriers (than the amplitude distributing one), which can then have larger powers (or amplitudes), for the same BER limit (with the sum of a larger number of sinusoids, the probability for clipping is smaller). For channels where all subcarriers are active for both power distributing LA and amplitude distributing LA, the performance is quite similar, and $A_{\text{eq}} \approx A_{\text{bud}}$ and $P_{\text{eq}} \approx P_{\text{bud}}$.

Figure 6.13 shows similar plots for the case of asymmetric clipping. The same discussion is valid as for results in Fig. 6.12.

6.2.3 Capacity estimate

To recapitulate, previous investigations showed achievable rates when square-QAM modulation formats are used, under the condition that a certain error rate is not to be exceeded. The overall best rate performance was seen by the conventional loading algorithm (introduced in the Section 6.1.2), where the total signal modulating power is chosen so that the background noise and clipping have an approximately same influence on the resultant error rate.

In conventional coherent systems, maximum rates obtained by a LA are upper-bounded by the system capacity obtained by the water-filling algorithm. Following figures present a similar comparison in the considered optical wireless system. Figure 6.14 a shows the rates already introduced in Fig. 6.12 a, for $\text{BER}_{\text{TOT}} \leq 2 \cdot 10^{-3}$ and $\text{BER}_{\text{TOT}} \leq 2 \cdot 10^{-5}$, in case of symmetric clipping (solid plots). The two dashed plots represent the capacity estimate obtained by classical water-filling, where for

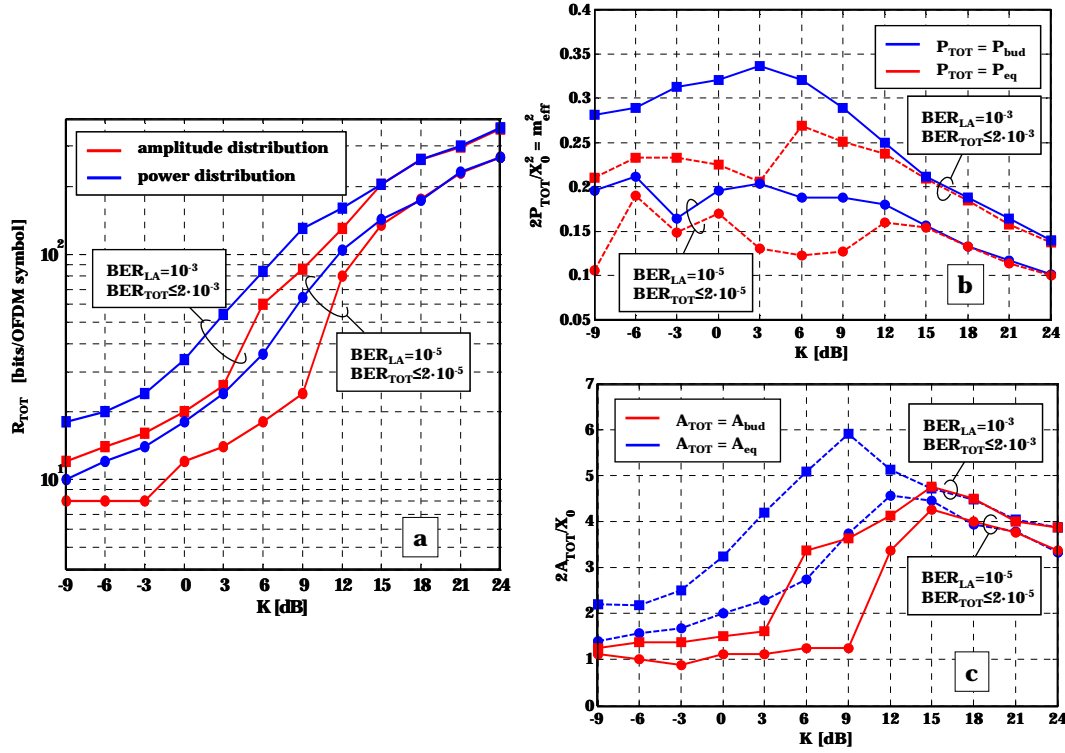


Figure 6.13: Comparison of the (a) achievable rates in terms of bits per OFDM symbol, (b) total signal modulating powers and (c) sums of amplitudes when loading algorithm with amplitude distribution (red plots) and loading algorithm with power distribution (blue plots) are implemented and asymmetric clipping is enforced. Assumed are $N = 64$, $P_O = 400$ mW and $S = 1$ W/A.

each channel realization (K -factor) the corresponding P_{bud} is taken, as found in Fig. 6.12 c. There are two capacity plots because permissible power budgets for $BER_{TOT} \leq 2 \cdot 10^{-3}$ are larger than for $BER_{TOT} \leq 2 \cdot 10^{-5}$.

Figure 6.14 a shows the capacity estimates significantly above the rate curves. The reason is that water-filling solution assumes a sum of Gaussian distributions, which superimpose in front of the optical source to result in a Gaussian distribution $\mathcal{N}(X_0, 2P_{bud})$. The ratios between the standard deviation and the mean value of this distribution (i.e., the effective modulation indices) are presented in Fig. 6.14 b (left y -axis). As symmetric clipping is considered (in 0 and $2X_0$), these ratios correspond to equivalent clipping probabilities of several permill to several percent (on both sides together). Once all the subcarriers are active, further increase of K -factor results in smaller modulating power increase with respect to DC (i.e., X_0), leading to decrease of the permissible clipping probabilities for the same overall BER limit. The water-filling algorithm provides optimistic capacity estimates (especially for lower K -factors) since it assumes transmission in a channel without

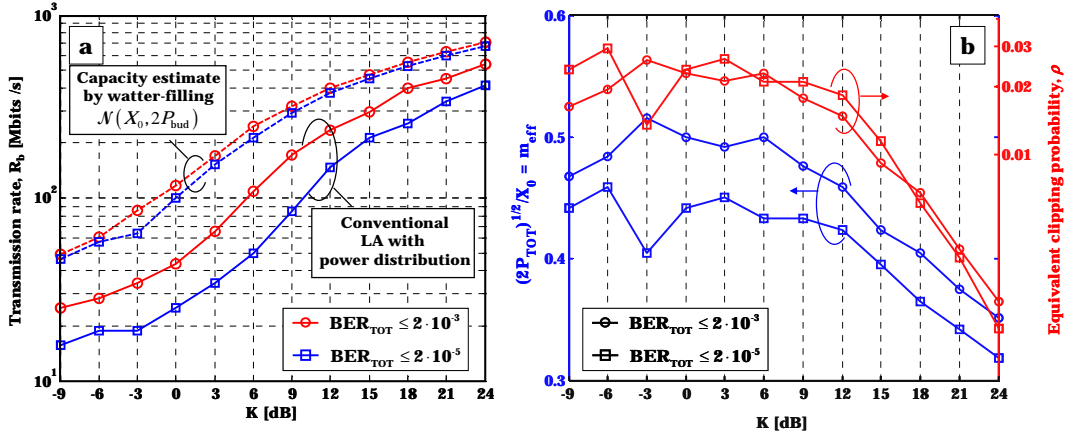


Figure 6.14: (a) Achievable rates against capacity estimate provided by water-filling for the same signal power, (b) The ratio between signal standard deviation and mean value (i.e., the effective modulation depth, left y-axis) and the corresponding probability for symmetric clipping (right y-axis) of the equivalent Gaussian distribution $\mathcal{N}(X_0, 2P_{bud})$ considered by the water-filling estimate. Assumed are $N = 64$, $P_O = 400$ mW and $S = 1$ W/A.

loss of energy through clipping. It can be said that water-filling results in an upper bound on the achievable rates of a system based on discrete square-QAM, which is tighter as the clipping probability decreases. On the other hand, when compared to the capacity upper bound obtained by the trigonometric moment space method (given in Chapter 5.2), one can conclude the achievable rates lie significantly below this bound (somewhere between the green and the blue plot in Fig. 5.13). This gap is an indication that the ensemble of several discrete square M-QAM modulations is not the capacity-achieving input for MSM systems with complex modulation. Nevertheless, from the practical point of view, water-filling estimate can be used to upper-bound the achievable rates of the OFDM-based system, when such relatively low clipping probabilities are considered.

Figure 6.15 shows corresponding plots for the case of asymmetric clipping. The same discussion is valid as for results in Fig. 6.14.

When figures for symmetric and asymmetric clipping are compared, it can be observed that the corresponding capacity plots are relatively close to one another. This is because the power budgets are often not significantly different for the two clipping cases. Also, it can be seen that corresponding rate plots (for symmetric and asymmetric clipping) are quite close to each other, as a result of the fact that an increase in the percentage of clipped samples does not linearly translate into the increase of errors. This is easily understandable since the OFDM symbols which are corrupted with errors from clipped samples on the negative side, usually

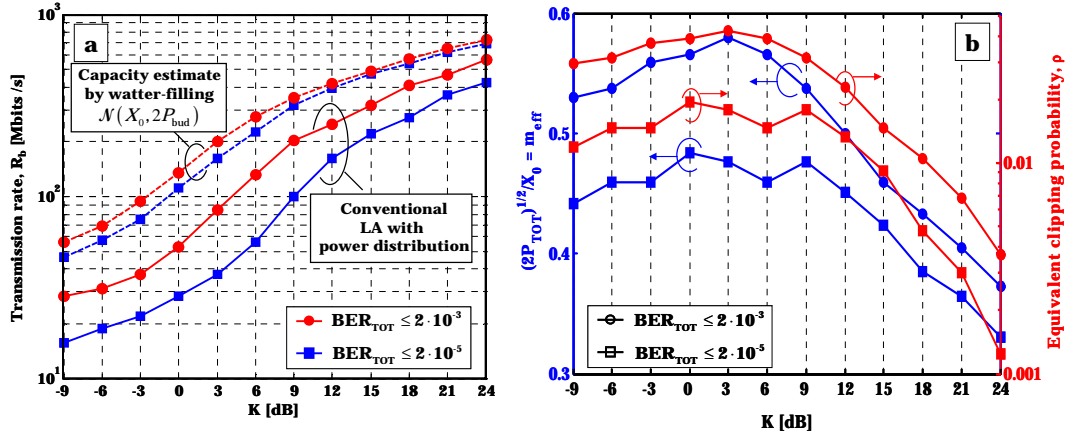


Figure 6.15: (a) Achievable rates against capacity estimate provided by water-filling for the same signal power, (b) The ratio between signal standard deviation and mean value (i.e., effective modulation index, left y -axis) and the corresponding probability for asymmetric clipping (right y -axis) of the equivalent Gaussian distribution $\mathcal{N}(X_0, 2P_{bud})$ considered by the water-filling estimate. Assumed are $N = 64$, $P_O = 400$ mW and $S = 1$ W/A.

produce errors on many subcarriers, so that enforcing clipping also on the positive side can not make an already erroneous symbol be even more false.¹⁴

When comparing the capacity estimates from Fig. 6.14 a and Fig. 6.15 a to the results in Fig. 5.13, it should be noted that these capacity plots take into account different clipping probabilities for different K -factors (according to Fig. 6.14 b and Fig. 6.15 b), whereas the plots in Fig. 5.13 consider a constant clipping probability over the whole K -range shown. Otherwise, the same (water-filling) algorithm is applied, only with different ρ parameters. Also note that the values for m_{eff} in the cases shown in Fig. 6.14 b and Fig. 6.15 b lie between 0.3 and 0.6, which is in agreement with the discussion made in the Chapter 4.2, regarding the influence of laser diode nonlinearity (recall that with 63 fundamentals at the laser input, the worst case C/I ratio is above 20 dB, for $m_{eff} \leq 0.6$).

6.2.4 Implementation of power control

Clearly, to maximize the transmission rate, the system must recognize the channel state and adaptively set the permissible P_{bud} which is then distributed over the subcarriers. Therefore, some form of *power control algorithm* is needed to provide the P_{bud} for each channel state. One possibility could be to use a look-up-table, which would consist of the maximum permissible P_{bud} values for all relevant channel

¹⁴This is true for the considered BERs, where the QAM symbol error still corresponds to the error on one bit.

realizations under a BER_{TOT} targets of interest. One example is shown in Fig. 6.16. Further on, each time when the loading algorithm is performed, it is done in 2 steps

1. A *reference loading* is performed, with a virtual budget fixed to a value $P_{\text{bud,ref}}$. The value of $P_{\text{bud,ref}}$ is chosen upfront so that the loading algorithm results in a different number of used subcarriers, or different total number of bits they carry for each channel realization of interest. That way, this virtual loading is an indicator of the present channel state (K -factor).
2. From the look-up-table the actual P_{bud} is extracted (for the K -factor implicitly found in step 1), and the loading algorithm is performed again, this time with a “real” P_{bud} to obtain the actual transmission parameters.

Such implementation does not depend on the number of subcarriers, and only requires doubling the time for the loading algorithm execution (since it needs to be performed twice), which is not an issue in the considered channel, as is discussed in the Section 6.3.

REFERENCE LA		LOOK-UP-TABLE (N=64, $P_o=400$ mW, S=1W/A)							
N=64 $\text{BER}_{\text{ref}} = 10^{-3}$ $P_{\text{bud,ref}} = 0.6$ W R_{TOT}		symmetric				asymmetric			
	K [dB]	$\text{BER}_{\text{TOT},1} \leq 2 \cdot 10^{-3}$ $P_{\text{bud}} [\text{W}] \cdot 10^3$...	$\text{BER}_{\text{TOT},n} \leq \text{xx}$ P_{bud}	...	$\text{BER}_{\text{TOT},1} \leq 2 \cdot 10^{-3}$ $P_{\text{bud}} [\text{W}] \cdot 10^3$...	$\text{BER}_{\text{TOT},n} \leq \text{xx}$ P_{bud}	
98	-10	18,75				23,75			
104	-9	17,5				22,5			
108	-8	16,875				21,25			
116	-7	20				25			
124	-6	18,75				23,125			
134	-5	17,5				21,25			
148	-4	19,375				23,75			
164	-3	21,25				25			
182	-2	21,875				24,375			
202	-1	21,25				25,625			
222	0	20				25,625			
242	1	20				25			
264	2	20,625				26,875			
288	3	22,5				28,75			
312	4	24,375				30,625			
336	5	26,25				32,5			
360	6	28,125				34,375			
384	7	30				36,25			
408	8	31,875				38,125			
432	9	33,75				40			
456	10	35,625				41,875			
480	11	37,5				43,75			
504	12	39,375				45,625			
528	13	41,25				47,5			
552	14	43,125				49,375			
576	15	45				51,25			
600	16	46,875				53,125			
630	17	48,75				55			
644	18	49,375				55,625			
660	19	50				56,25			
678	20	50,625				56,875			
702	21	51,25				57,5			
730	22	51,875				58,125			
756	23	52,5				58,75			
	24	53,125				59,375			
	25	53,75				60			

Figure 6.16: Illustration of a possible power control implementation.

6.2.5 Summary of main results

Finally, Fig. 6.17 presents the improvement of transmission rates which can be achieved by designing the system adaptively (i.e., implementing a loading algorithm) and by implementing a power control mechanism which takes controlled clipping into account.

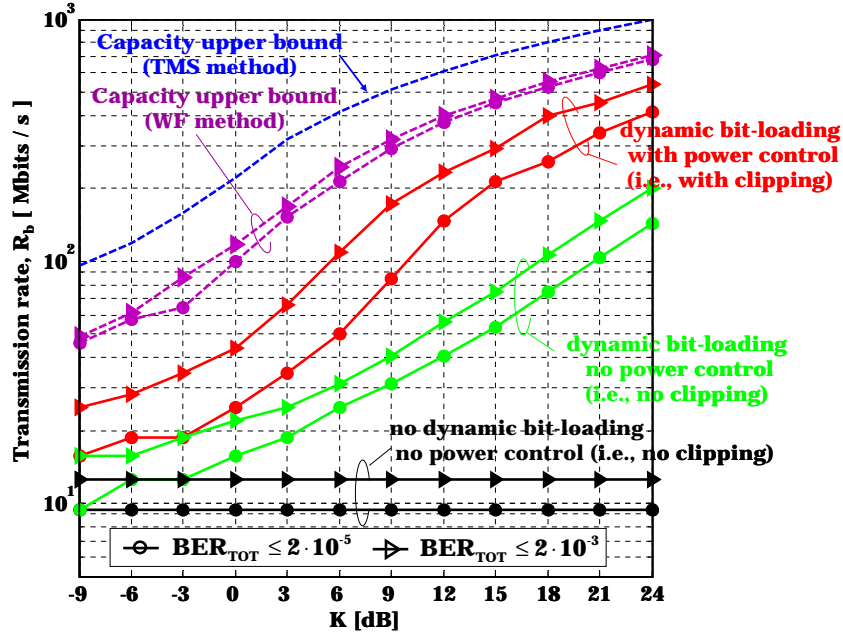


Figure 6.17: Comparison of the achievable rates in a considered system when no adaptation, adaptation only with loading algorithm, or adaptation with both loading algorithm and power control is implemented. Assumed are $N = 64$, $P_O = 400$ mW and $S = 1$ W/A.

Results are shown in terms of transmission rate in [bits/s]. Rate curves of the same color reflect the cases when $\text{BER}_{\text{TOT}} \leq 2 \cdot 10^{-3}$ (triangle marker) and $\text{BER}_{\text{TOT}} \leq 2 \cdot 10^{-5}$ (circle marker) are considered. It can be seen how beneficial it is to use the channel information and dynamically adapt the transmission rate, with respect to the system without such a feature, where the necessary link reliability (BER) limits the system throughput to the worst-case scenario (i.e., purely diffuse channel). Even under most conservative clipping conditions (no clipping permitted), for many channel realizations from the range of interest, achievable rates increase substantially. By further adding a power control mechanism (which allows to achieve approximately the same BERs for over the whole K -factor range of interest, by permitting different degrees of clipping depending on the channel state), controlled clipping can be allowed in the system, which results in significant further transmission rate improvements, but is also the simplest method for power efficiency enhancement. This power control mechanism allows the system to dynamically adapt the clipping probabilities at the optical source based on

the channel state. As a reference marker, the upper bound for system capacity obtained by the trigonometric moment space method and the estimates of the capacity by water-filling are included in the figure. All rates R_b of practical systems are obtained by equation (6.10), which does not account for transmission overhead. Overhead includes effects of guard interval (cyclic prefix) length, training sequences, error correction coding etc., and decreases the net transmission rate, depending on the implementation design. Note that by combining the information from Fig. 6.17 with the information regarding the distribution of K -factors in the room from Fig. 2.12 (depending on transmitter radiation index, Tx-Rx distance and Rx orientation), achievable rates throughout the room can be derived, showing that under many channel conditions encountered in the considered communication scenario significant improvements in the rate performance are possible when the adaptive modulation technique is deployed.

6.3 System coherence interval

Proper data recovery requires knowledge of channel influence on the transmitted signal. Received signal on each orthogonal subcarrier is influenced by a certain channel coefficient (sub-bands with flat fading). The estimation at the Rx can be therefore performed subcarrier-wise, using training sequences. Estimated values of channel coefficients are used for equalization of the transmitted data symbols, as illustrated in Fig. 6.18. The time interval in which the channel estimation (and the needed length of the training sequence) should be repeated depends on the variation of the channel coefficients H_n in time.

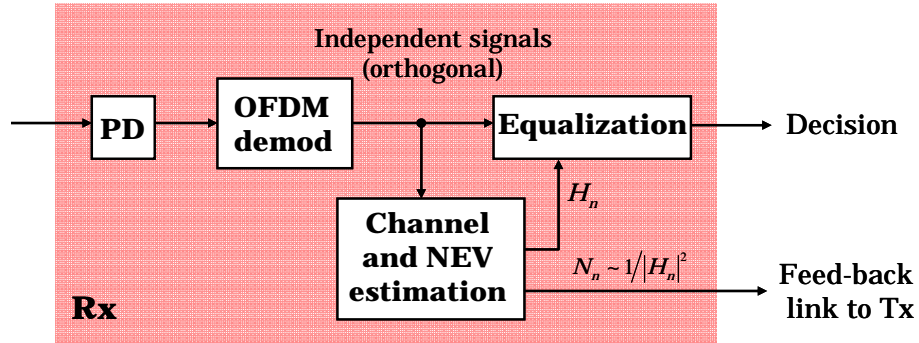


Figure 6.18: Functional blocks at the Rx dependent on the system coherence interval.

Capacity analysis in Chapter 5, as well as loading schemes presented in Section 6.1, assume that some form of channel information is made available to the transmitter via a low-speed feed-back link. For a dynamically adaptive system considered in this work, it is, therefore, necessary to determine also how often the loading information needs to be updated. The loading algorithm is based on the noise enhancement over the subcarriers ($N_n \sim 1/|H_n|^2$), so the update interval depends

on the variation of $|H_n|^2$. This information, referred to as the *noise enhancement vector* (NEV, vector of N_n coefficients for all subcarriers), is sent back to the Tx side,¹⁵ where it is fed in the loading algorithm. The measure of the needed system update interval is usually referred to as the *system coherence interval*, SCI. This coherence interval indirectly determines the needed transmission rate (bandwidth) of the feed-back link.

For a fixed configuration (Tx, Rx, and objects in the room), the OW channel can be thought of as fixed. The channel changes, however, if the configuration changes (as it was introduced in Chapter 2.3). This variation in space can be translated into a variation in time, if (at least) one of the channel determining components is considered to be moving at a certain speed. In this thesis, a simple approach is adopted - a receiver is assumed to be moving across the room at a walking speed $v \approx 1.4$ m/s. The channel variation is investigated by ray tracing simulations and compared to the already introduced analytical model ((2.6), (2.7)). For the simulations, the Tx is considered fixed at the center of the ceiling pointing downwards in the model room (Fig. 2.1), and the Rx is considered moving at the desk-top plane always with the orientation upwards. Given the symmetry of the assumed configuration, the largest channel variation occurs when the receiver shifts¹⁶ along the diagonal of the desk-top surface, as illustrated in Fig. 6.19 a. In this way, the channel response is obtained in 35 points on the diagonal (from center to the corner in 10 cm steps), and for each position (x), channel coefficients are calculated in 63 equidistant frequencies up to 100 MHz (subcarrier frequencies $f_1 = 100 \text{ MHz}/64$, $f_n = n f_1$, $n = 2..63$), thus forming a grid (matrix) of channel coefficients as in Fig. 6.19 b, to serve as the basis for the investigations to follow. Using this simple approach, the variation of the OW channel is coarsely estimated, and consequences on the feed-back link design are discussed.

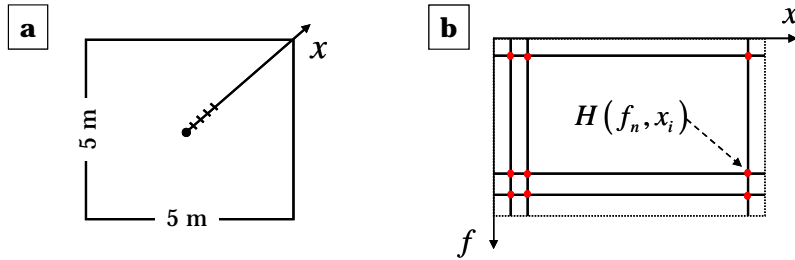


Figure 6.19: (a) Illustration of the scenario for simulations, (b) Grid of frequencies and Rx positions for which channel coefficients are obtained via ray tracing simulations.

¹⁵This feed-back link is usually incorporated in the uplink connection, as will be further discussed later on.

¹⁶For simplicity sake, tilting of the receiver is not considered.

6.3.1 Variation in the OW channel

Diffuse (DIFF) component

The diffuse (DIFF) component of the OW channel (resulting from diffuse reflections off the room surfaces) is under investigation first. The model describes this component with a low pass function (recall (2.7))

$$H(f, x) = \eta_{\text{DIFF}} \frac{e^{-j2\pi f \Delta T(x)}}{1 + j \frac{f}{f_0}} = \frac{\eta_{\text{DIFF}}}{\sqrt{1 + \left(\frac{f}{f_0}\right)^2}} e^{-j \left[\arctan\left(\frac{f}{f_0}\right) + 2\pi f \Delta T(x) \right]}, \quad (6.34)$$

where the additional phase shift $2\pi f \Delta T(x)$ results from the delay of the diffuse signal relative to the LOS signal arrival at the receiver. According to the model, the signal amplitude remains constant at one frequency regardless of the receiver position x , while the phase changes due to different moments of the signal arrival. Variation of the diffuse channel coefficient at frequency f , and between two Rx positions x_1 and x_2 can be expressed by the phase shift due to the delay difference

$$H(f, x_2) = H(f, x_1) e^{-j2\pi f [\Delta T(x_2) - \Delta T(x_1)]}. \quad (6.35)$$

Figure 6.20 presents an example of the DIFF channel impulse response (exponential according to the simple analytical model, recall (2.6)) for two extreme points on the diagonal - in the center and near the corner of the room, with the time axis origin at moment of signal transmission (the LOS signal is omitted). Clearly, with Rx in the center of the room, it takes longer for the reflected signals to arrive. The longer the paths the more the response is attenuated with respect to the case when the Rx is positioned near the walls¹⁷.

In the model, the delay ΔT represents the time that passes before the signal from the first reflection arrives at the receiver. Nevertheless, since it can be seen from Fig. 6.20 that the responses are not ideally exponential¹⁸, we consider defining ΔT as the delay between the LOS and the mean of the diffuse impulse response.

Figure 6.21 a presents the amplitude of the diffuse frequency response with Rx at several positions on the desk-top diagonal (on y -axis, $10 \log$ of the amplitude is presented). It can be observed that the amplitude follows the behavior predicted by the model (2.7)(red plot in the figure) very well up to the cut-off frequency (cut-off frequency according to the model and according to the ray tracing results is found to be between 8-9 MHz). The best agreement between simulation results

¹⁷The attenuation is also a consequence of the chosen Tx-Rx constellation. With Rx in the center and facing the ceiling, strong first reflections are not caught.

¹⁸The largest discrepancies occur in the beginning of the impulse response, when the first few reflections arrive, which are responsible for response on higher frequencies.

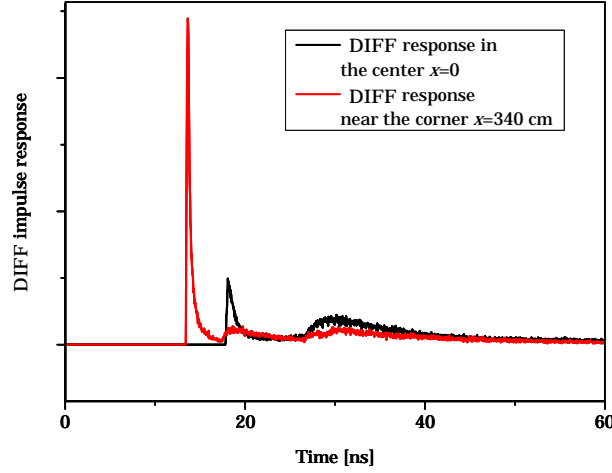


Figure 6.20: Ray tracing simulation of the diffuse channel impulse response for Rx placed in the center (black) and near the corner (red) of the room.

and the model can be seen when the receiver is placed at the center of the desk-top surface, because this scenario resembles a case of an integrating sphere the most. In the considered f - x -grid, this range is covered by the first 6 subcarriers. Figure 6.21 b presents the behavior of the diffuse response amplitude of the first dozen subcarriers when the receiver is moved from the center to the corner of the room. This figure confirms that at low frequencies (approximately up to f_0), the diffuse channel amplitude remains practically constant throughout the room as predicted by the model. At the higher frequencies the response can not be accurately predicted (by the assumed model) due to the already mentioned discrepancies of the impulse response to the exponential fall around the first few reflections.

When moving across the diagonal, a frequency dependent phase shift is expected, due to the factor $2\pi f \Delta T(x)$ in (6.34). Figure 6.22 a shows the variation of the first 12 coefficients of the diffuse channel (corresponding to the first 12 subcarrier frequencies) when the receiver is moved from the center (vector tips denoted by blue diamonds) to the corner (vector tips denoted by red circles) of the room.

Ray tracing simulations confirm the behavior predicted by the model. There is basically no amplitude change on a particular frequency - all vector tips remain on the same circles, and frequency dependent phase shifts are also obvious. In Fig. 6.22 b, the phase difference between the corner and center receiver positions is shown for different subcarriers. The expected linear dependence (due to the delay ΔT) is well confirmed up to f_0 (the first 6 subcarriers) and it can be explained by the difference in the global phase delay $\Delta T_g = \Delta T(x_{\text{corner}}) - \Delta T(x_{\text{center}})$. Extrapolating from the first 6 simulation results $\Delta T_g \approx 9$ ns, whereas when determined from the impulse response graph (Fig. 6.20) the difference in the delays between first reflections is

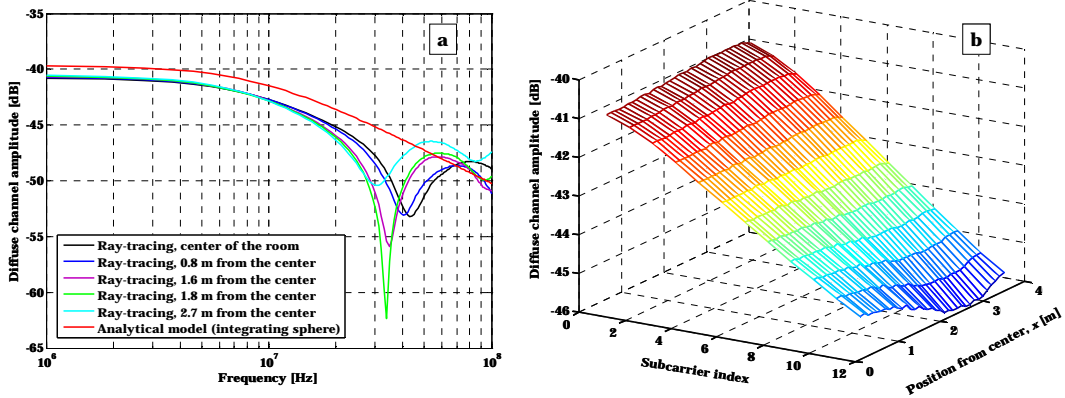


Figure 6.21: (a) Behavior of the diffuse frequency response amplitude with Rx placed in the center of the room - comparison of the simulated results to the analytical model (6.34) (or (2.7)), (b) Behavior of the diffuse frequency response amplitude for the first dozen subcarriers ($f_1 = 100/64 = 1.5625$ MHz and $f_{12} = 12f_1 = 18.75$ MHz), when the receiver is being shifted from the center to the corner of the room. Assumed are: $\langle \rho \rangle = 0.6$, Rx shifts along the diagonal, other parameters as in Table A.2.

4.5 ns and between mean responses is 14 ns. Therefore, the order of magnitude is correct. Better accuracy can hardly be expected with the simple model assumed.

LOS component

Figure 6.23 presents the two LOS impulse responses obtained by ray tracing with the receiver placed at the center and in the corner of the room.

It can be seen that the amplitude of LOS signal diminishes as the Tx-Rx distance increases, as well as that the signal needs $\Delta T_g = \Delta T(x_{\text{corner}}) - \Delta T(x_{\text{center}}) = 6.62$ ns longer to arrive to the receiver near the corner of the room. According to the model, it is expected that the LOS component varies with receiver position as

$$H(f, x) = \eta_{\text{LOS}}(x) e^{-j2\pi f T_{\text{LOS}}(x)}, \quad (6.36)$$

where $T_{\text{LOS}}(x)$ represents the time needed for the signal to arrive at the receiver via direct path, as illustrated in the inset of Fig. 6.23. Hence, at one frequency, both the amplitude (2.10) and the phase are expected to change with position, due to attenuation and different times of signal arrival. When $\phi(x) = \psi(x)$ is assumed in (2.10), the amplitude variation can be expressed as

$$\eta_{\text{LOS}}(x) = \eta_{\text{LOS},0} \frac{1}{\left(1 + \frac{x^2}{h'^2}\right)^{\frac{m+3}{2}}} = \eta_{\text{LOS},0} \frac{1}{\left(1 + \frac{x^2}{h'^2}\right)^2}, \quad (6.37)$$

where the second equality is valid for Lambert index $m = 1$ (assumed in the simulations as well). The LOS amplitude is not frequency dependent, and Fig. 6.24

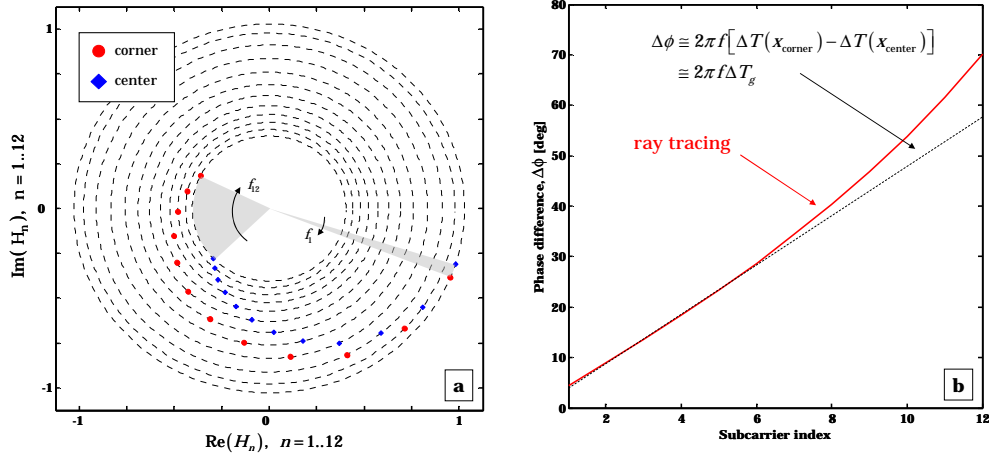


Figure 6.22: (a) Variation of the diffuse frequency response for the first dozen subcarriers ($f_1 = 100/64 = 1.5625$ MHz and $f_{12} = 12f_1 = 18.75$ MHz) and (b) Variation of the diffuse frequency response phase, when the receiver is being shifted from the center to the corner of the room.

shows that it perfectly fits the mathematical model. In (6.37), $h' = r(x = 0)$ denotes the shortest Tx-Rx distance, when Rx is placed directly under Tx.

The same can be concluded from Fig. 6.25, where the variation of the LOS channel coefficients is presented for several subcarriers, when the receiver is moved from the center towards the corner of the room. Between any two positions, the variation of the amplitude is the same (two dashed circles) regardless of the frequency, whereas the phase variation increases with the subcarrier frequency.

Figure 6.26 shows the dependence of the coefficient phase on the frequency (a) and on the receiver position (b), resulting from ray tracing simulations. The frequency dependence is perfectly linear which indicates that the phase variation between any two points can be attributed to the delay between the signal arrivals. The derivation of the total delay between the two extreme positions

$$\Delta T_{\text{LOS},g} = \frac{1}{2\pi N} \sum_{n=1}^N \frac{|\Delta\phi_n|}{f_n} \quad (6.38)$$

renders approximately 6.46 ns, which is very close to the value of 6.62 ns, obtained directly from the impulse responses (Fig. 6.23). Figure 6.26 b shows a nonlinear relationship between the phase and position, which is caused by the nonlinear relationship of the time delay and position

$$T_{\text{LOS}}(x) = T_{\text{LOS},0} \sqrt{1 + \frac{x^2}{h'^2}}, \quad (6.39)$$

due to the chosen geometry (see the inset of Fig. 6.23).

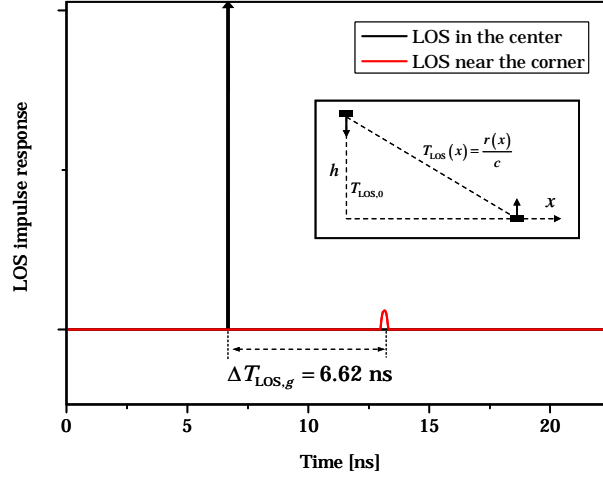


Figure 6.23: Impulse response of LOS channel for Rx placed in the center (black) and near the corner (red) of the room. Inset: T_{LOS} is the time signal needs to arrive at the receiver via direct path, $T_{\text{LOS},0} = T_{\text{LOS}}(x = 0)$, c is the speed of light.

6.3.2 Quantification of channel and SNR deviation

Correlation approach

The system coherence interval (measure of the needed system update) can be estimated based on the correlation of the varying parameters. Namely, a criterion could be set in a way that it is necessary to update the system parameters (i.e., estimate the channel and/or perform the loading algorithm anew) if they change by a certain degree. As mentioned before, variation of channel coefficients $H(f, x)$ is of importance for the channel estimation, whereas the update of the loading algorithm is determined by the variation of $|H(f, x)|^2$.

For one frequency, (complex) channel coefficients are obtained for $M = 35$ receiver positions on the diagonal (denoted by x , see Fig. 6.19 a)

$$H_i = H(x = x_i) \in \mathbb{C}, \quad i = 1, \dots, M. \quad (6.40)$$

If a relative measure of change between any two observation points x_i and x_k is denoted by $\beta_{ik}^{\text{CE}} \in \mathbb{R}$, the criterium for new channel estimation is given by

$$H_{i+k} = (1 + \beta_{ik}^{\text{CE}})H_i, \quad (6.41)$$

where

$$\beta_{ik}^{\text{CE}} = \frac{H_{i+k} - H_i}{H_i} = \frac{H_i^* H_{i+k} - |H_i|^2}{|H_i|^2}, \quad (6.42)$$

and for a new loading by

$$|H_{i+k}|^2 = (1 + \beta_{ik}^{\text{LA}})|H_i|^2, \quad (6.43)$$

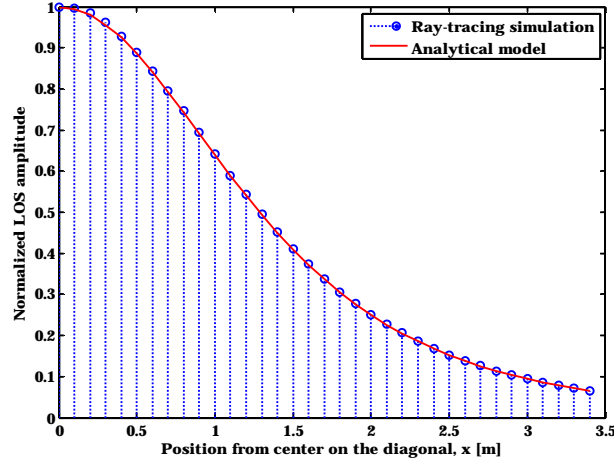


Figure 6.24: Behavior of the LOS frequency response amplitude when the Rx is moved from the center of the room ($x = 0$ m) to the corner of the room ($x = 3.5$ m) on the desk-top surface diagonal - comparison of the simulated results to the analytical model (6.37). Assumed $h' = 2$ m.

where

$$\beta_{ik}^{\text{LA}} = \frac{|H_{i+k}|^2 - |H_i|^2}{|H_i|^2}. \quad (6.44)$$

For both criteria, corresponding factors β_{ik} are determined in all observation points, in order to build a mean relative measure of change when the receiver is relocated k positions away $\beta_k^{\text{CE(LA)}} \in \mathbb{R}$

$$\beta_k^{\text{CE(LA)}} = \frac{1}{M-k} \sum_{i=1}^{M-k} \beta_{ik}^{\text{CE(LA)}}. \quad (6.45)$$

Results are presented in Fig. 6.27 as a plot of *correlation coefficients* $\rho_k = 1 - |\beta_k|$ (for both types of β_k). Also, the Rx position is scaled by walking speed of 1.4 m/s, so to obtain time on the x -axes. Plots reflect the channel and SNR variations in time, for the most critical subcarriers (subcarriers experiencing the largest change), since the estimation of the update interval needs to be based on the most conservative parameters. In the case of the diffuse channel, only the first 6 subcarriers were considered (within the 3 dB bandwidth).

It can be observed that the diffuse channel is practically static and that the overall channel variation is mostly influenced by the change of LOS signal alone. Moreover, the LOS signal variation is mostly a result of its amplitude variation, so the SNR changes faster than the channel, and therefore it is the critical factor for both the estimation and loading update intervals.

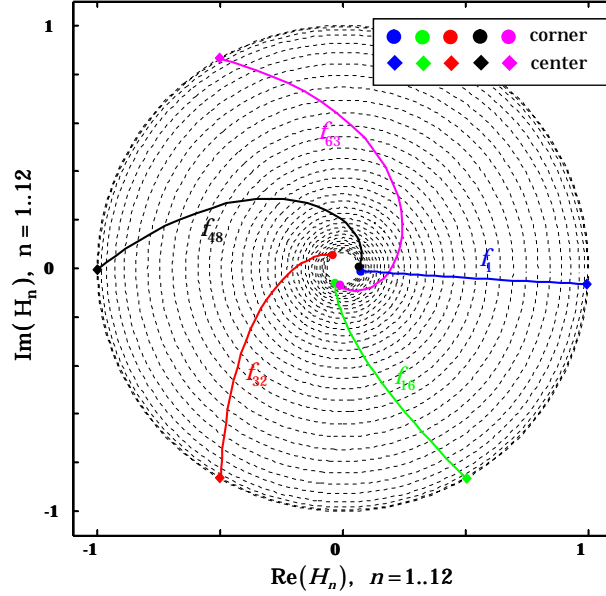


Figure 6.25: Variation of the LOS frequency response (for particular subcarriers 1, 16, 32, 48 and 63) when the receiver is being shifted from the center to the corner of the room. Assumed $N = 64$ (63 usable subcarriers) in $B = 100$ MHz and $h' = 2$ m.

As a sufficient criterium one may require the system update when the SNR correlation coefficient falls under $\rho = 0.75$.¹⁹ From the graph it can be estimated that new information would be needed every 150 ms. This is quite relaxed compared to the radio systems. Because of the low subcarrier frequencies (up to 100 MHz) and low speeds (walking speed, since indoor environment is considered) Doppler is comparable to FM radio but not to typical mobile systems at 900 MHz, 2 or 5 GHz.

Interference approach

While the correlation analysis gives a good rule-of-thumb result, it provides no information on error-vector magnitude degradation. Therefore, when deployment of high order modulation formats is foreseen, a more practical value for system design results from the approach where the channel deviation is modeled as an interference, separately from the AWGN. When noise is disregarded, the output signal on each subcarrier at the position (moment) x is given by

$$\begin{aligned} Y_n(x) &= H_n(x)X_n \\ &= H_n(x_0)X_n + E_n(x - x_0)X_n, \end{aligned} \quad (6.46)$$

¹⁹In general, it is absolutely necessary to update the system parameters when the SNR changes 3 dB ($\rho = 0.5$), since the difference between the needed SNR for adjacent M-QAM formats ($M = 4, 8, 16, 32, \dots$) is approximately 3 dB. When only square M-QAM formats are considered, this difference is approximately 6 dB, which relaxes the needed ρ to 0.25.

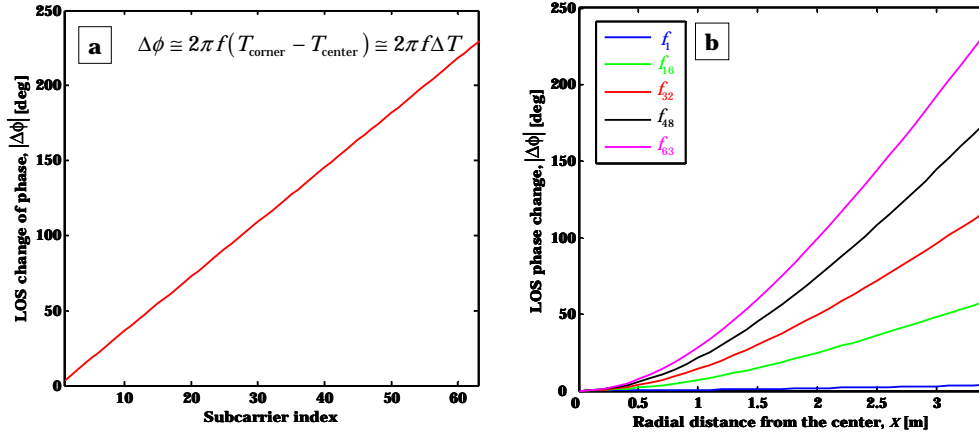


Figure 6.26: Variation of the LOS frequency response phase as a function of (a) frequency and (b) receiver position.

where $x = x_0$ is the moment of the last, presumably perfect, channel estimation and $E_n(x - x_0) = H_n(x) - H_n(x_0)$ is the error of the channel coefficient until the position (moment) x . Channel deviation since last estimation can then be modeled as interference

$$\text{SIR}_n(\Delta) = \frac{\langle |H_n(x_0)|^2 \rangle}{\langle |H_n(x_0 + \Delta) - H_n(x_0)|^2 \rangle}, \quad (6.47)$$

where $x = x_0 + \Delta$. The signal-to-interference ratio (SIR) is determined as the mean value over all receiver positions Δ apart. The result of this approach is presented in Fig. 6.28 for the first and the last subcarrier. Clearly, as the receiver moves away since the last estimation (i.e., as time passes), the error grows and SIR decreases. Moreover, the greatest error is present on the highest frequency because of the largest phase variation. The horizontal reference lines relate to the SNR (in the AWGN channel) needed for reception with BER of 10^{-6} .

Given that the SIR is determined without taking into account the noise, the results can be interpreted as following: if 64-QAM on the last subcarrier (worst case) needs to be detected with $\text{BER}=10^{-6}$, it is necessary to update the system at least every 30 ms, which is clearly much shorter period then estimated by the correlation approach. For a practical system, even this figure should be reduced if the margin for noise and other effects is included. If a larger BER is allowed in the system (since, e.g., error correction coding is foreseen), the resulting system coherence interval is more relaxed. From Fig. 6.28 the update interval would almost double when $\text{BER}=10^{-3}$ is considered (dashed line). Even with this approach the obtained estimate for system coherence interval is relaxed compared to Long Term Evolution (LTE) systems where typical times are in the range of 1 ms for the channel and about 5 ms for the SIR.

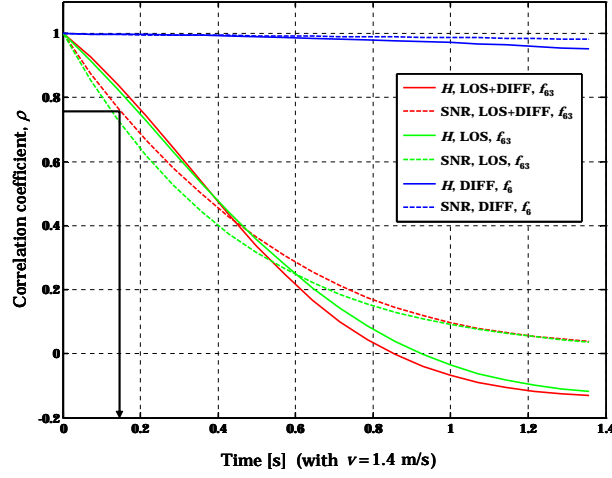


Figure 6.27: Correlation approach characterization of the system coherence interval.

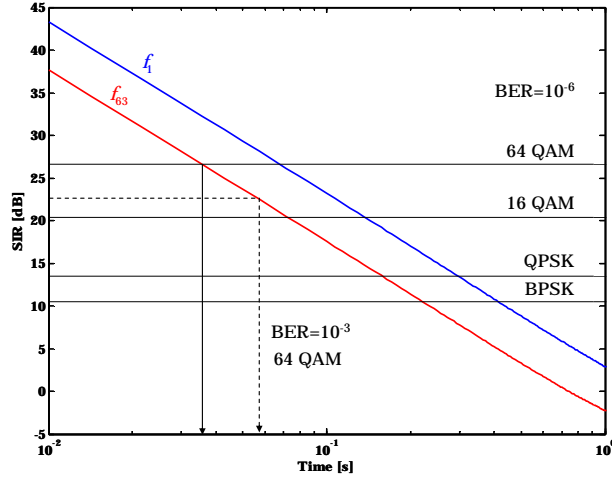


Figure 6.28: Characterization of the system coherence interval by modeling the channel variation as interference. Assumed indoor speed $v = 1.4$ m/s.

6.3.3 Consequences on link design

With the system coherence interval estimated at 30 ms (interference approach), link overhead is discussed in the following text.

Figure 6.29 provides a plausible timing chart of communication between the two units (T1 and T2) on a physical layer. T1 sends a frame consisting of a training sequence (consisting of N_T OFDM symbols), control fields (containing modulation and coding information of the data), channel quality information (containing the noise enhancement vector (NEV) information) and data. Training sequence is used at T2 to perform the channel estimation and to acquire the information for the

feed back link. Channel coefficient on subcarrier n is obtained by averaging over all training symbols $H_n = \langle Y_n(i)/X_n(i) \rangle$, for $i = 1, \dots, N_T$, so the training sequence needs to be sufficiently long to average out the Gaussian noise well. It should be sent often enough to keep the system always up-to-date, yet it is desirable to keep the overhead low.

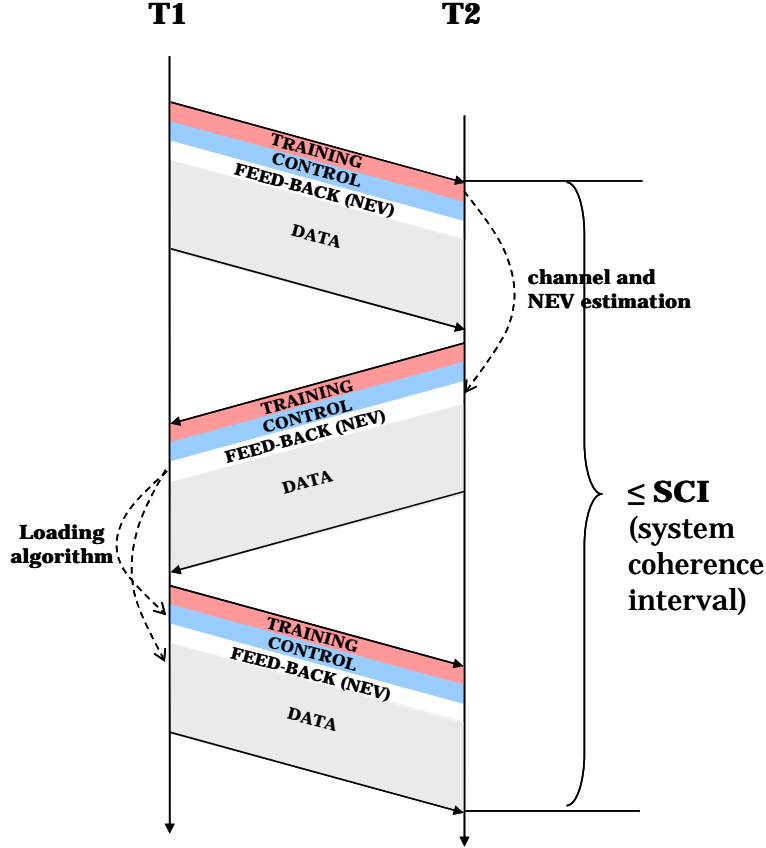


Figure 6.29: Timing chart of communication between the transmitter and receiver unit on a physical layer.

The estimated channel coefficients and the training sequence are used to evaluate the noise enhancement coefficients on the subcarriers, which are sent back as the feed-back field within the next frame from T2 to T1. Upon arrival at T1, NEV is used in the loading algorithm to provide the information regarding modulation format and power on each subcarrier, which is then imprinted on the next data transmission block (from T1 to T2). This control information is also sent to T2 (control field), so that T2 knows how to demodulate the data on different subcarriers. It is necessary that the system parameters remain unchanged from the instant when the training is sent until the corresponding data block arrives at the T2 (shown in Fig. 6.29 as system coherence interval).

Regarding the short indoor distances (few meters), propagation delays are ~ 10 ns, and can be neglected at this point (since an OFDM symbol lasts $T_{\text{OFDM}} = (2N + L)/2B = 0.72 \mu\text{s}$, with $N = 64$, $L = 16$ and $B = 100$ MHz assumed). The processing delays are also negligible since presumably everything is implemented on an field programmable gate array (FPGA) device, typically used for trials, or on an application specific integrated circuit (ASIC), used for large numbers of units. If we neglect the propagation delay and the times needed for signal processing, then approximately the length of 3 frames should fit within the system coherence interval

$$3(T_{\text{T}} + T_{\text{CNT}} + T_{\text{FB}} + T_{\text{D}}) \leq \text{SCI}, \quad (6.48)$$

where T_{T} , T_{CNT} , T_{FB} and T_{D} represent the transmission delay for training sequence, control block, feed-back and data block, respectively. These frame sections can be assumed to last N_{T} , N_{CNT} , N_{FB} and N_{D} OFDM symbols, respectively, so

$$(N_{\text{T}} + N_{\text{CNT}} + N_{\text{FB}} + N_{\text{D}}) T_{\text{OFDM}} \leq \frac{\text{SCI}}{3}. \quad (6.49)$$

Note that the illustrated case corresponds to the general situation when the channel from T1 to T2 is not the same as the channel from T2 to T1. Otherwise, T2 could use the estimated channel information to immediately perform the loading algorithm and imprint the results on data transmitted from T2 to T1. In this case, only two frames need to fit within the SCI.

With OFDM symbol period of $T_{\text{OFDM}} = 0.72 \mu\text{s}$, a SCI of 30 ms allows approximately 13800 OFDM symbols within a transmission frame. A training sequence can be assumed to be a few tens of OFDM symbols long (as can be deduced from [103]), and its influence can therefore be neglected. It is also realistic to consider that the durations of control and feed-back sequence have the same order of magnitude ($N_{\text{FB}} \approx N_{\text{CNT}}$), which leaves

$$(N_{\text{D}} + 2N_{\text{FB}}) \leq \frac{\text{SCI}}{3T_{\text{OFDM}}} \approx 13800 \quad (6.50)$$

OFDM symbols in a frame. Since all information is digitally processed, for each noise enhancement coefficient, a 12-bit representation can be assumed (common in FPGA), meaning that $12 \times 63 = 756$ bits need to be transmitted within one T_{FB} (and approximately the same in T_{CNT}). Since these frame sections carry control information, their transmission needs to be extremely reliable. This is why we can assume that the bits are distributed only over the first few (2-3) subcarriers which are always usable for transmission due to the diffuse link. Moreover, for the same reason, powerful correction coding should be assumed. If we consider loading the first 3 subcarriers with 2 bits each, the total of 1500 bits (for control and feed-back)

can be transmitted within $1500/6 = 250$ OFDM symbols, and the double amount of information (due to correction coding) within 500 OFDM symbols. In this case the data block of each OFDM frame can consist of

$$N_D \leq \frac{\text{SCI}}{3T_{\text{OFDM}}} - 2N_{\text{FB}} \approx 13800 - 500 = 13300 \quad (6.51)$$

OFDM symbols. Therefore, due to the generous system coherence interval, the length of frame overhead for control and feed-back information, can be as low as 2-4%.

Chapter 7

Summary and Outlook

This thesis provides an evaluation of potentials of the adaptive modulation technique in an OW OFDM-based system for typical WLAN applications. Even though the regarded technique is known from DSL systems and is well researched for RF systems, specific properties of the optical wireless channel called for a separate assessment of its performance and link layer design.

Deployment of adaptive OFDM in an OW system brings about various benefits, among which are providing a large spatial coverage with very simple and inexpensive optical front-ends without the need for pointing and tracking mechanisms, or generation of diversity; support of user mobility in combination with a reliable connectivity; efficient completely digitalized signal processing which allows for the simplest linear equalization at the receiver even in a multipath channel. Differently than in radio systems, because of the real-valued channel (direct detection), an OW system requires generation of a baseband OFDM signal, which forces doubling the size of FFT blocks and faster DAC/ADC. Nevertheless, because of the fact that relatively moderate number of subcarriers is allowed for good spectral efficiency and that the maximum signal bandwidth is limited from the photodetector side, implementation with standard devices is not challenging. Another issue which is different to the RF OFDM systems concerns synchronization. Even though it has not been considered within this thesis, direct detection allows much simpler link synchronization. In OW OFDM systems, only the frame start needs to be located, whereas in RF systems care about the carrier frequency offset must be taken as well.

As one of the first, if not the first work to recognize the attractiveness of modulation adaptive technique in optical wireless systems, this thesis handles two main questions: *Where are the theoretical limits for transmission rate?* (Chapter 5), and *How and to what extent can these theoretical limits be approached in practice?* (Chapter 6). In the context of assessing the practicability of the investigated approach, supporting questions such as: *What is the adequate channel model and which set of model parameters are realistic and suitable for the considered commu-*

nication scenario? (Chapter 2), *How should the simple optical front-ends of the system be designed to allow broadband transmission with sufficient power budget?* (Chapter 4), *How often do system transmission parameters need to be updated in order to retain reliable connectivity?* (Chapter 6) are also thoroughly dealt with. In the text that follows, a summary of main conclusions is made.

At the start of investigations, a communication scenario has been defined representing a bidirectional SISO link for a typical moderate-size room and a simple suitable channel model was adopted and adapted to the scenario. It was shown that in the considered space both bandwidth and magnitude of the communication channel strongly vary depending on the Tx-Rx configuration, indicating the attractiveness of the adaptive modulation approach.

Opto-electrical front-ends

Opto-electrical front-ends for the considered communication scenario were analyzed and their relevant components and parameters discussed, in order achieve a design as simple as possible, which at the same time allows broadband transmission with enough link budget.

A significant issue of the permissible optical power at the transmitter was investigated in accordance with the safety regulations standard. Eye and skin safety were both regarded. It was found that the highest power is permissible for an extended Lambertian source operating in the 850 nm window. The more critical tissue is the skin because the specified maximum permissible exposure level for the skin must be measured at the position where the exposure level is the highest, i.e., when the skin is in direct contact with the emitter output window. A value of 400 mW was taken as the limit at Tx (assuming that a minimal size of fully illuminated surface is enforced by optic, i.e., that the apparent source diameter is extended to 10 mm).

The influence of non-linearity in the transmission system was investigated. Because of wireless transmission and direct modulation of the LD by analogue signals at frequencies up to 100 MHz, imperfect linearity of the static laser optical-power-vs-current characteristic was assumed responsible for distortions. Whereas for passband analogue transmission (in radio or fibre-optics), the subcarrier range can be chosen so that only the third-order intermodulation products fall in range of fundamental tones, baseband OFDM transmission in this work required consideration of second order distortions as well. The goal of investigations was to determine the permitted power of the signal modulating a typical commercially available optical source component (which could be used in the system considered), under the condition that the distortions result in a certain minimum C/I ratio. In that sense, a model characteristic of a typical 850 nm VCSEL was experimentally determined and used to evaluate the influence of non-linearities when an arbitrary number of

equidistant tones is present at the laser input. Investigations have shown that the most influential type of distortion are the second order intermodulation products, and that the highest impairments occur at the lowest fundamental frequencies. With a constant signal modulating power, the maximum influence of the second order intermodulation products (i.e., at the lowest tone frequency) becomes practically independent on the increasing number of tones. As a consequence, the maximum modulating signal power is determined rather by the acceptable C/I ratio, than by the number of tones at the source input. With the condition on C/I ratio fulfilled, the influence of nonlinearities can be neglected and the system transmission rate is primarily limited by the safety constraint and the errors which occur due to the noise or clipping of high peaks, if such is permitted at the optical source. For improvement of the link budget, a laser array is proposed as a suitable source.

At the receiver side, due to the non-directed link considered in this work, a non-imaging hemispherical optical concentrator was chosen because of its simplicity combined with a large field-of-view (FOV). Such concentrator also exhibits a relatively constant gain for a wide range of angles of incidence. With a planar photodetector, and a desired FOV of 180° , the concentrator gain is limited to about 3 (depending on the glass refraction index). A hemispherical thin-film optical filter placed on top of the hemispherical optical concentrator is considered in this thesis since, with proper dimensioning of concentrator radius, it is theoretically possible to chose the filter bandwidth arbitrarily small, while exhibiting negligible influence of the incoming light ray incidence angle. Filter bandwidth and concentrator radius are however inversely proportional to each other, which means that, especially for portable terminals, size and weight requirements can limit the concentrator radius.

The investigation of the photodetector, as the main component of the OW receiver, considered both silicon PIN and APD photodiode type, as well as low-impedance and transimpedance preamplifiers in combination with detectors of different sizes. Influence of both transient time (due to photodiode type) and RC -constant (due to preamplifier type) on the receiver bandwidth was discussed. Besides bandwidth, the receiver sensitivity was investigated based on conducted noise analysis. The detector of choice in this work was an APD plus a TIA preamplifier with a typical value of feed-back resistor. Beside the advantage in sensitivity (even though not a great one) over a PIN with the same TIA, such receiver is shot-noise limited and the link analysis could be performed with the assumption of a white gaussian noise in the channel (as is usual for optical wireless links). On the other side, a corresponding PIN receiver is dominated by the frequency dependent noise. Moreover, for the same reason, the performance of an APD receiver is not as strongly dependent on the particular preamplifier design as the performance of a PIN one. Investigation of the APD-TIA receiver bandwidth revealed that for small detector sizes, the transient-time component determines the total bandwidth (due to large

optimal multiplication factors), whereas, in the case of large detectors, the RC -bandwidth becomes the deciding factor. It was shown that the desired bandwidth of 100 MHz can be realized with a detector whose diameter is about 3 mm. Moreover, with a TIA, the open-loop amplifier gain presents another degree of freedom, which can be used to fine-tune the design to the aimed bandwidth.

The desired bandwidth of 100 MHz resulted in a choice of detector size and preamplifier type. It was clear that whereas the baseband signal bandwidth for poor channel states is limited by the OW channel itself, for good channel states it is actually limited by the bandwidth of the photodetector. With a photodetector radius decided on, the suitable concentrator radius, and the bandwidth of the optical filter were chosen. The discussion of the opto-electrical front-ends was not intended to optimize the receiver design, but aimed to provide a set of realistic receiver parameters useful for further investigations.

Capacity

One of the fundamental questions discussed in this thesis is the capacity of OW OFDM systems, as the theoretical limit for the performance. A trigonometric moment space (TMS) method to upper-bound the capacity of flat optical IM/DD channels with complex multiple subcarrier modulation was generalized to the case of a frequency-selective channel. According to this method, the rate-maximizing transmission bandwidth (i.e., optimal number of used subcarriers when a subcarrier bandwidth is fixed) is determined. Although the upper bound is derived only under the assumption of a high SNR, it was shown that the SNRs of the active subcarriers are in the required range. Theoretical potentials of an adaptive system were recognized by comparing the obtained upper bound to an upper bound of the statically designed (non-adaptive) system. Namely, for many channel states which can be expected in the considered scenario, channel knowledge at Tx allows much higher transmission rates to be realized. The benefits over the non-adaptive system grow with the channel K -factor (with dominance of the LOS over the diffuse signal contribution). In addition, it was found that the capacity grows with the logarithm of the square of the safety constraint, and that the subcarrier bandwidth has practically no influence on the capacity as long as the total system (channel) bandwidth is kept constant.

Furthermore, upper-bounding the capacity according to another, conservative constraint (which considered all subcarriers in phase) was also performed, due to its relevance for the practical system without clipping. It was seen that choosing such straightforward but conservative constraint leads to theoretical rates which are much below the theoretical performance bound of the system under the true constraint. The straightforward approach (to satisfy the constraints in a practical system) would lead to a performance with significant room for improvement. Using

an example of a two-dimensional space, these different bounds for the input signal were illustrated, showing that the TMS bound is not tight for very small number of subcarriers. It is however known that the bound becomes tighter with the number of dimensions (subcarriers).

Since the amplitude variation of an OFDM signal can be approximated by a Gaussian distribution, in the case when the true constraints on the transmit signal in the IM/DD channel can be translated into a power constraint, the conventional water-filling could become relevant for a practical system.¹ Because of this, compromises and relaxations of the initial problem, required to properly apply the water-filling solution in an optical IM/DD system, were explained, and the resulting estimate of system capacity was discussed. Since a Gaussian distribution at the channel input results in imminent clipping, it was concluded that the water-filling solution can only provide an upper bound (since it does not foresee any loss of signal energy) of the OFDM system with the same power at the channel input. There is a trade-off between the tightness of this upper bound and signal power (i.e., capacity). This estimate is too pessimistic when the clipped-Gaussian is well approximated by a Gaussian one, indicating that the Gaussian distribution at the input of the regarded channel is not capacity achieving (it lies well below the TMS plot). Naturally, for the same signal constraints, much higher power can be chosen so that the obtained capacity estimate lands in vicinity of the TMS method. However, at the same time the actual distribution entering the optical channel does not resemble Gaussian so well, and the approximative result of the water-filling process is increasingly overestimated.

Signal processing

Since the chosen design of opto-electrical front-ends allows a baseband OFDM signal bandwidth of 100 MHz, consideration of typical channel lengths and desire of low information redundancy by guard band insertion, resulted in the choice for the size of FFT/IFFT blocks, i.e., maximal number of 63 independent subcarriers (64 including DC). With the design of relevant system parameters completed, potentials of adaptive transmission were investigated, with a goal to maximize the total link throughput, at the same time efficiently eliminating the effects of a dispersive environment.

Throughput maximization in a practical system, considering different orders of square M-QAM modulation, and complying to a certain maximal BER, is obtained by deployment of loading algorithms. Based on the system constraints and channel state information, such algorithms vary the transmitted power level and constellation size of the signals on individual subcarriers. An optimal loading al-

¹Since this is not the case in general, only an estimate of capacity can be obtained.

gorithm from DSL systems was investigated in this work. Since the OW channel is not power-limited in a conventional manner, it was found that this algorithm cannot be applied in its original form, since the constraints on the signal in front of the optical source are not sufficient for the electrical AC signal power to be uniquely determined. Instead, the conservative constraint regarding the sum of subcarrier amplitudes was considered (the straightforward practical way to satisfy the true constraints on the signal), and the algorithm was adjusted to perform optimally under these conditions. As expected, it was shown that the achievable rates are always below the capacity bound derived under such conservative constraint.

A comparison to the capacity upper bound derived by the TMS method and considering the true constraint on the transmit signal, showed that assuming the signals on all subcarriers in phase (which is theoretically a worst-case scenario) leads to transmission rates which are well below the actual theoretical limits (even when being aware that the TMS upper bound is not everywhere tight). The conservative bound is of significance when signal non-negativity is desired to be upfront eliminated, without additional measures.

In this work, the influence of controlled clipping in front of the optical source on the performance of the OFDM-based OW system was investigated, in terms of both error and transmission rate. Hereby, both symmetric and asymmetric clipping of the digital time samples was considered.

The effect of clipping on the performance was investigated by means of MC simulations. Allowing for clipping means that the modulating power of the transmission signal can be increased. Two approaches were regarded and compared - either to directly increase the total signal power, or to increase it indirectly by increasing the sum of subcarrier amplitudes (based on the already investigated conservative approach), in both cases until the error performance reaches a certain threshold.

Firstly, effects of clipping onto BER were qualitatively evaluated and described. Investigations have shown that clipping introduces a random shift of the sent symbol depending on amplitudes and phases of the symbols sent on all subcarriers, and that the variance of clipping errors for each subcarrier can be modeled as Gaussian distributed interference. Predicting the influence of the clipping effect and incorporating it into the loading algorithm was however found difficult, since clipping depends on data sent on all subcarriers, which can then no longer be considered as independent. Simulation results have also shown that for many channel states of interest additional errors due to clipping become relevant for the error performance only when the amplitude (or power) budget is significantly increased. It was shown that for different channel states different budgets are permitted with the same condition on error performance. The most increase over the conservative budget was shown possible for channel states where additional amplitude (power) budget was

used to activate additional subcarriers, which in turn reduced the possibility for clipping.

The investigations of the possible rate enhancements when controlled clipping is tolerated led to conclusion that it is recommendable to target a system error performance with approximately equal contributions from noise in the channel and clipping at the optical source. Comparison of the two approaches to allow signal clipping (i.e., increase of the amplitude budget and distribution of amplitudes according to the modified loading algorithm, and increase of the power budget and distribution of powers according to the conventional loading algorithm) resulted in better performance of the conventional approach. The reason is that once clipping was allowed in the system, apart from maximal DC value, the limit on BER presented the only constraint, which was indirectly the constraint on the signal power (rather than amplitude). The amplitude distributing algorithm achieved lower rates because it considered an additional constraint on the sum of amplitudes. As already mentioned, such algorithm would be relevant if the performance without clipping is forced, or the crest-factor and amount of clipping are of interest.

As in conventional power-limited systems, upper bound for the rates achieved in the considered OW system by the power distributing algorithm (without the constraint on the sum of amplitudes) is also given by the water-filling solution. Signal power budgets were taken as the variance of the Gaussian distributed inputs, and equivalent clipping probabilities were determined for different channel states. Based on these equivalent clipping probabilities (not exceeding a few percent), it was concluded that the approximation of the input signal with a Gaussian distribution was acceptable for the considered cases, but with more clipping allowed, this upper bound becomes less tight. On the other hand, when the rates achieved by the loading algorithm were compared to the capacity upper bound obtained by the TMS method, it was seen that they lied significantly below this bound, meaning that the ensemble of several discrete square M-QAM modulations was not the capacity achieving input for MSM systems with complex modulation. Nevertheless, from the practical point of view, since the water-filling bound lies below the TSM bound when relatively low equivalent clipping probabilities are considered, this estimate can be used for upper bounding the OFDM-based system capacity.

Finally, it can be concluded that, under many channel conditions encountered in the considered communication scenario, great improvements in the performance are possible when the adaptive modulation technique is deployed, and especially when power control (clipping) is included. It was shown that for the channel states where the LOS component is free and equally strong as the diffuse channel component ($K \approx 0$), the transmission rate can be improved about several times compared to a non-adaptive system without clipping; for a LOS component 10 times stronger than the diffuse one, improvements in rate are about one order of

magnitude (even 20-fold); and for really good channels with the LOS component 20 times above the diffuse, about 40 times higher rates are achievable. In this way, with realistic parameters of the considered system, gross rates between a couple of tens and several hundred Mbit/s are shown to be possible, depending on the channel state. When the rates achieved in case of symmetric and asymmetric clipping were compared, even though they were never lower for the asymmetric case, it was observed that they were relatively close to each other, as a result of the fact that doubling the percentage of clipped samples does not linearly translate into an increase of errors. Namely, the OFDM symbols, corrupted with errors from clipped samples on the negative side, usually already produced errors on many subcarriers, so that additional clipping of the positive peaks in the same symbols had no great extra impact. Furthermore, when power control is implemented in the system, i.e., controlled clipping permitted, the system power-efficiency is also significantly improved. For implementation of the power control in a practical system, an approach based on a look-up-table was proposed. The suggested implementation does not depend on the number of subcarriers, and only requires doubling the time for the loading algorithm execution.

System update interval and feed-back link design

Investigations were made to determine the needed interval for system update regarding the channel state information and the loading information. Whereas the frequentness of channel estimation depends on the variation of the channel coefficients on the subcarriers H_n in time, loading algorithm update depends on the variation of $|H_n|^2$. For a fixed Tx-Rx configuration, the channel is invariant, however, it varies, if the configuration changes. In this work, variation in space was translated into variation in time, by assuming that the receiver unit was moving across the room at walking speed. Variation of both LOS and diffuse channel components was investigated by ray tracing simulations and compared to the adopted analytical model. Two approaches were considered to quantify the channel variation, i.e., to estimate the system coherence (update) interval. From the correlation approach it was obvious that the diffuse channel component is practically static and that the channel variation is mostly influenced by the change of LOS component. Moreover, since the LOS variation is mostly a reflection of its amplitude variation, the SNR ($|H_n|^2$) changes faster than the channel (H_n), presenting therefore the deciding factor for both the channel estimation and loading update intervals. While the correlation analysis gave a good rule-of-thumb result, it was unable to provide information on error-vector magnitude degradation. Since high order modulation formats are foreseen in the system, a more practical value of the system coherence interval for system design was obtained by the approach where the channel deviation was modeled as interference (interference approach). A value of 30 ms, resulting from this approach when 64-QAM is foreseen to be detected with bit-error ratio of 10^{-6} , is much more relaxed compared to mobile radio sys-

tems. This estimate was then used to coarsely derive some consequences on the feed-back link design (required speed and bandwidth). Because the channel varies very slowly, transmission of the feed-back link information can be imbedded within first few subcarriers (always suitable for transmission due to the diffuse channel). Even in the case of significant rate increase due to powerful correction coding, the overhead of the transmission frame (consisting of the feedback and other control information) is only a few percent. Note also that having a fixed placement of the feed-back channel greatly simplifies the channel monitoring function.

Outlook

The investigations conducted in this thesis considered performance of the adaptive modulation technique assuming the simplest possible optical front-ends in a SISO link. Such approach was taken in order to be able to recognize the contributions of the chosen transmission method. It can be imagined that further improvements in rate performance could be obtained if additionally optical front ends were optimized, either in a way to be able to lift the safety limit, or to dimension the photodetector better (to obtain larger photodetectors with enhanced bandwidth, or to achieve better receiver sensitivity). Moreover, angle or space diversity could be realized in the system as well as the true multiple-input-multiple-output approach. The price would of course be the higher system complexity (of front-ends but also of signal processing).

In this work only adaptive modulation was considered, and it was not foreseen that errors introduced by clipping are corrected. Adaptive signal coding or ways to correct the clipping errors could also be investigated, in order to enhance the error performance. If the clipping errors could be corrected, further improvements in power efficiency but also transmission rate could be expected. There is, however, a trade-off with the involved overhead (e.g., if some subcarriers need to be reserved as pilot tones for the correction mechanism). Moreover, implementation in hardware and experimental verification of the obtained results are points for future work.

The capacity discussions made in this thesis are valid not only for optical wireless MSM systems, but also for fibre-optic IM/DD systems with multiple subcarrier modulation. A certainly very interesting question would be to investigate modulating formats (input distributions) which would better approach the capacity bound obtained by the TMS method.

Appendix A

Parameters for Ray Tracing Simulations

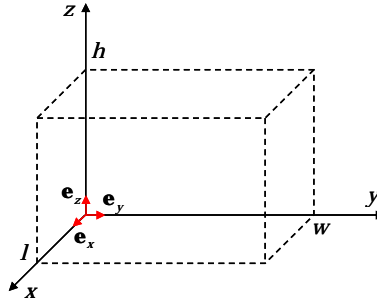


Figure A.1: Coordinate system for the model room from Fig. 2.1.

Parameter	$(l, w, h) = (3, 3, 3) \text{ m}$	$(l, w, h) = (5, 5, 3) \text{ m}$	$(l, w, h) = (8, 8, 3) \text{ m}$
$\langle \rho \rangle$	0.6		
Tx position (x, y, z)	$(1.5, 1, 0) \text{ m}$	$(2.5, 2, 0) \text{ m}$	$(4, 2.5, 0) \text{ m}$
Tx orientation (e_x, e_y, e_z)	$(0, 0, 1)$		
Tx radiation index m	1		
Rx position (x, y, z)	$(1.5, 2, 0) \text{ m}$	$(2.5, 4, 0) \text{ m}$	$(4, 5.5, 0) \text{ m}$
Rx orientation (e_x, e_y, e_z)	$(0, 0, 1)$		
Rx FOV _{det}	π		
Rx surface	*78.5 cm ²		

Table A.1: Parameters for simulations regarding the influence of the model room size (Fig. 2.6). * For ray tracing simulations an unrealistically large detector surface is assumed in order to decrease the simulation time.

Parameter	Value
Room dimensions (l, w, h)	$(5, 5, 3)$ m
Tx position (x, y, z)	$(2.5, 2.5, 2.9)$ m
Tx orientation (e_x, e_y, e_z)	$(0, 0, -1)$
Tx radiation index m	1
Rx position (x, y, z)	A $(2.5, 2.5, 0.9)$ m B $(3.35, 3.35, 0.9)$ m C $(4.2, 4.2, 0.9)$ m D $(4.95, 4.95, 0.9)$ m
Rx orientation (e_x, e_y, e_z)	$(0, 0, 1)$
Rx FOV _{det}	π
Rx surface	*78.5 cm ²
$\langle \rho \rangle$	0.6

Table A.2: Parameters for investigation of diffuse component delay (Fig. 2.9). * For ray tracing simulations an unrealistically large detector surface is assumed in order to decrease the simulation time.

Further remarks:

In order to eliminate the possibility for a direct path, and yet to have the possibility of the first (strongest) reflections, both Tx and Rx were placed on the floor pointed directly to the ceiling.

For simulations regarding investigation of the influence of average reflectivity of room surfaces (Fig. 2.7) as well as receiver FOV (Fig. 2.8), all fixed parameters were taken from the column corresponding to the $5 \times 5 \times 3$ m model room.

The diffuse channel gain obtained with such parameter is also unrealistically large. The realistic gain can be obtained by simple scaling

$$\eta_{\text{DIFF,actual}} = \eta_{\text{DIFF}} \frac{A_{\text{Rx}}}{78.5 \text{ cm}^2}$$

when a detector of realistic surface A_{Rx} is assumed.

Appendix B

Elimination of ISI with OFDM

This appendix provides a detailed mathematical description of how an OFDM system inherently deals with dispersion effects in multipath channels. The relevant relationships between signals at different points of transmission chain are illustrated in Fig. B.1.

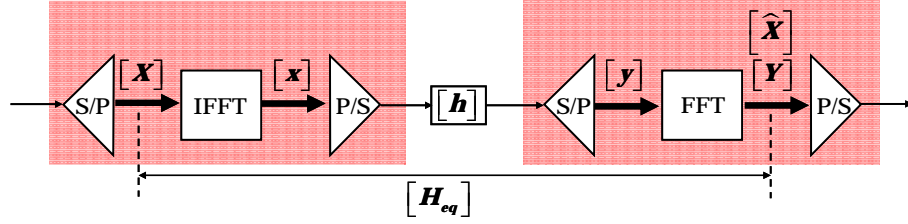


Figure B.1: DFT-based block-wise transmission over an LTI channel.

As in Fig. B.1, let $[\mathbf{X}]$ be the input vector of size $N \times 1$ with complex symbols as elements and let $[\mathbf{x}]$ be the output of the IFFT block, which passes through an LTI channel with discrete impulse response h . Also, let the signal at the channel output be $[\mathbf{y}]$, and the result of OFDM demodulation at the receiver after FFT, be $[\mathbf{Y}]$. The relationships between the signals are summarized as

$$\begin{aligned} [\mathbf{x}] &= \text{IFFT}\{[\mathbf{X}]\}, \\ y = x * h &\Leftrightarrow [\mathbf{y}] = [\mathbf{h}][\mathbf{x}], \\ [\mathbf{Y}] &= \text{FFT}\{[\mathbf{y}]\}, \end{aligned} \quad (\text{B.1})$$

where matrix $[\mathbf{h}]$ is introduced in (3.3). $\text{FFT}\{\cdot\}$ and $\text{IFFT}\{\cdot\}$ denote the discrete (i.e., fast) Fourier transform operators. Definition of FFT/IFFT operation pair

$$\begin{aligned} \text{FFT} : \quad X_n &= \sum_{k=0}^{N-1} x_k e^{-j2\pi nk/N}, \quad n = 0, \dots, N-1 \\ \text{IFFT} : \quad x_k &= \frac{1}{N} \sum_{n=0}^{N-1} X_n e^{j2\pi nk/N}, \quad k = 0, \dots, N-1 \end{aligned} \quad (\text{B.2})$$

is used in this thesis. Even with a linear convolution in the physical channel, because of DFT operators before and after the channel, system input-output relationship between $[\mathbf{Y}]$ and $[\mathbf{X}]$ (in the frequency domain) is not linear and the equalization at the receiver can not be simple. To be able to deploy the simplest equalization, which is linear frequency domain correction by division

$$[\widehat{\mathbf{X}}] = [\mathbf{H}_{\text{eq}}]^{-1} [\mathbf{Y}], \quad (\text{B.3})$$

output-input relationship in frequency domain needs to be linear, that is, a circular convolution is necessary in time domain [56]. The channel matrix $[\mathbf{h}]$ needs hence to be modified in such way that the linear convolution, which actually takes place in the channel, appears as a circular convolution.

If $[\mathbf{h}]$ is modified into $[\mathbf{h}_{\text{pc}}]$ so that in time domain the linear convolution with $[\mathbf{h}_{\text{pc}}]$

$$y = h_{\text{pc}} * x \Leftrightarrow [\mathbf{y}] = [\mathbf{h}_{\text{pc}}] [\mathbf{x}], \quad (\text{B.4})$$

appears as circular with $[\mathbf{h}]$, in frequency domain this corresponds to the multiplication of their DFTs, i.e.,

$$\begin{aligned} \text{FFT} \{[\mathbf{y}]\} &= \text{FFT} \{[\mathbf{h}]\} \text{FFT} \{[\mathbf{x}]\}, \\ [\mathbf{Y}] &= \text{FFT} \{[\mathbf{h}]\} [\mathbf{X}]. \end{aligned} \quad (\text{B.5})$$

From (B.5), it is clear how a circular convolution in the channel renders a linear input-output relationship (in f -domain).

The way to achieve circular convolution in the channel is to introduce a guard band between OFDM blocks in a form of cyclic prefix, which not only eliminates inter-block interference, but does exactly that what is needed for linear equalization. It namely, renders a diagonal equivalent channel matrix between $[\mathbf{Y}]$ and $[\mathbf{X}]$, and hence preserves the orthogonality among subcarriers, i.e., eliminates the inter-carrier-interference. A cyclic prefix of length L is inserted to the signal vector of length N by left matrix multiplication of the input vector with a typical circulant generator matrix

$$[\mathbf{G}_{\text{CP}}] = \begin{bmatrix} \mathbf{0}_{L \times (N-L)} & \mathbf{I}_L \\ \mathbf{I}_N & \end{bmatrix}. \quad (\text{B.6})$$

This matrix copies the last L samples of each OFDM symbol to its beginning, as illustrated in Fig. 3.6 b. Since they are corrupted by the channel, yet carry no information, the first L elements of each received block are simply discarded at the receiver before the FFT block by the left matrix multiplication with

$$[\mathbf{R}_{\text{CP}}] = [\mathbf{0}_{N \times L} \quad \mathbf{I}_N]. \quad (\text{B.7})$$

After the cyclic prefix is included and then removed, the equation (3.4) becomes

$$[\mathbf{R}_{\text{CP}}][\mathbf{y}(i)] = [\mathbf{R}_{\text{CP}}][\mathbf{h}_0][\mathbf{G}_{\text{CP}}][\mathbf{x}(i)] + [\mathbf{R}_{\text{CP}}][\mathbf{h}_1][\mathbf{G}_{\text{CP}}][\mathbf{x}(i-1)]. \quad (\text{B.8})$$

Part of the second summand $[\mathbf{h}_1][\mathbf{G}_{\text{CP}}][\mathbf{x}(i-1)]$ represents an $(N+L) \times 1$ vector, where only the first L elements could be different than 0. Thus, the removal of cyclic prefix eliminates exactly these L elements, annulling this summand completely and inter-block-interference with it. After discarding the cyclic prefix, the vector at the channel output (now N samples long) is

$$[\mathbf{y}(i)] = [\mathbf{R}_{\text{CP}}][\mathbf{h}_0][\mathbf{G}_{\text{CP}}][\mathbf{x}(i)]. \quad (\text{B.9})$$

Hence, by comparison to (B.4), the channel matrix has been modified into

$$[\mathbf{h}_{\text{pc}}] = [\mathbf{R}_{\text{CP}}][\mathbf{h}_0][\mathbf{G}_{\text{CP}}]. \quad (\text{B.10})$$

which is an $N \times N$ *pseudo-circulant* matrix¹

$$[\mathbf{h}_{\text{pc}}] = \begin{bmatrix} h(0) & 0 & \cdots & 0 & h(L_C - 1) & \cdots & h(1) \\ \vdots & \ddots & \ddots & & \ddots & \ddots & \vdots \\ h(L_C - 1) & \cdots & h(0) & 0 & \cdots & 0 & h(L_C - 1) \\ 0 & h(L_C - 1) & \cdots & h(0) & 0 & \vdots & 0 \\ \vdots & 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ & & & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & h(L_C - 1) & \cdots & \cdots & h(0) \end{bmatrix}.$$

System linearity is achieved thanks to an important property of any $N \times N$ circulant matrix, which is that it can be *diagonalized* by pre- and post multiplication with N -point FFT and IFFT matrices. These are defined as Vandermonde matrices [105]

$$[\mathbf{W}_N] = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & W_N & W_N^2 & \cdots & W_N^{N-1} \\ 1 & W_N^2 & W_N^4 & \cdots & W_N^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & W_N^{N-1} & W_N^{2(N-1)} & \cdots & W_N^{(N-1)(N-1)} \end{bmatrix}, \quad (\text{B.11})$$

$$[\mathbf{W}_N]^{-1} = \frac{1}{N} [\mathbf{W}_N]^*, \quad (\text{B.12})$$

respectively. Operator $[\cdot]^*$ denotes conjugation of the matrix elements. Elements of these matrices $W_N^{nk} = e^{-j2\pi nk/N}$ represent the so called *twiddle factors*, i.e points on a unit circle in a complex plane, with an equidistant angle spacing of $2\pi/N$. Figure B.2 shows the positions of twiddle factors in a complex plane for the case of $N = 8$.

¹An $N \times N$ circulant matrix \mathbf{C} of a sequence $[c_1, c_2, \dots, c_N]$ has its (k, l) element given by $c(k, l) = c_{(k-l) \bmod N}$, where $k, l = 1, 2, \dots, N$ [65], [104].

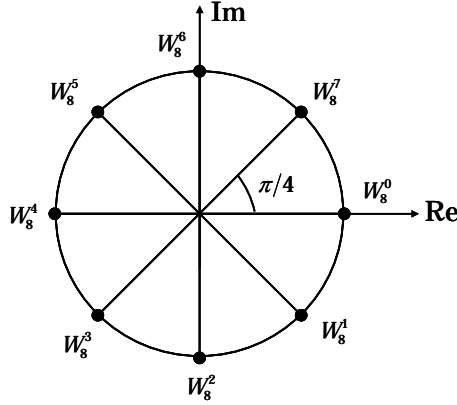


Figure B.2: Positions of twiddle factors in a complex plane for $N = 8$.

The equivalent channel matrix (between $[\mathbf{X}]$ and $[\mathbf{Y}]$) can hence be defined as

$$[\mathbf{H}_{\text{eq}}] = [\mathbf{W}_N] [\mathbf{h}_{\text{pc}}] \frac{[\mathbf{W}_N]^*}{N}. \quad (\text{B.13})$$

This resultant matrix $[\mathbf{H}_{\text{eq}}]$ is diagonal, with elements corresponding to Fourier coefficients of the circulant matrix. Since, in this case, $[\mathbf{h}_{\text{pc}}]$ represents a circulant convolution matrix of the channel impulse response, elements on the diagonal of $[\mathbf{H}_{\text{eq}}]$ are the channel frequency response on subcarrier frequencies, that is

$$H_{\text{eq}}(n, n) = H(n) = H(e^{j2\pi f_n}) = \sum_{k=0}^{L_C-1} h(k) e^{(-j2\pi \frac{nk}{N})}, \quad n = 0, \dots, N-1. \quad (\text{B.14})$$

The system O/I relationship is therefore $[\mathbf{Y}] = [\mathbf{H}_{\text{eq}}] [\mathbf{X}]$, i.e.,

$$\begin{bmatrix} Y(0) \\ Y(1) \\ \vdots \\ Y(N-1) \end{bmatrix} = \begin{bmatrix} H(0) & 0 & \cdots & 0 \\ 0 & H(1) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & H(N-1) \end{bmatrix} \begin{bmatrix} X(0) \\ X(1) \\ \vdots \\ X(N-1) \end{bmatrix}. \quad (\text{B.15})$$

so that each element of the output vector depends only on the corresponding element of the input vector as $Y(n) = H(n)X(n)$, which means that the orthogonality among subcarriers remains preserved and that the simplest equalization is possible. The channel equalization by left multiplication with the inverse equivalent channel matrix

$$[\widehat{\mathbf{X}}] = [\mathbf{H}_{\text{eq}}]^{-1} [\mathbf{Y}] = \frac{[\mathbf{W}_N]^*}{N} [\mathbf{h}_{\text{pc}}]^{-1} [\mathbf{W}_N] \quad (\text{B.16})$$

requires that the pseudo-circulant matrix has an inverse matrix, or equivalently that the channel transfer function has no zeros on the FFT grid ($H(n) \neq 0$, $(\forall n) \ n = 0 \dots, N-1$).

Appendix C

Radiation Safety Considerations

C.1 Maximum permissible Tx power based on AELs for eye safety

Case 1a. $\lambda = 900$ nm, time base $t = 3 \cdot 10^4$ s, collimated beam¹

Generally for near IR wavelengths, the angle of incidence α is obtained by placing the optical source (of aperture diameter a) at the distance r in front of the circle-shaped detector with aperture diameter d , as shown in the Fig. C.1. According to the norm, the values r and d depend on the wavelength range and the beam characteristic of the source in question.

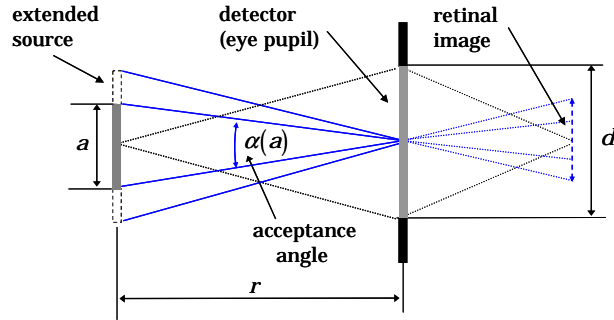


Figure C.1: Geometry of the measurement setup for calculation of the AEL values.

The apparent viewing angle α is dependent on the apparent source size, which is set with a circle shaped aperture stop of a variable diameter a . Then we can write

$$\alpha[\text{rad}] = 2 \arctan \frac{a}{2r} \approx \frac{a[\text{mm}]}{r[\text{mm}]} \quad (\text{C.1})$$

¹A “parallel” beam of radiation with very small angular divergence or convergence.

For larger (divergent) sources, the image on the retina grows with α , and the injury threshold in terms of the total permissible source power increases. Nevertheless, there are limits set on the viewing angle. The lower limit defined by $\alpha_{\min} = 1.5$ mrad (corresponding to image diameter $25 \mu\text{m}$) presents the apparent angle of incidence under which the source can be considered as point source. In this case, the retinal image is assumed to be a diffraction-limited spot of some $10 - 20 \mu\text{m}$, regardless of the source size. The upper limit defined by $\alpha_{\max} = 100$ mrad (corresponding to image diameter 1.7 mm), beyond which there is no spot-sized dependence of the retinal thermal injury threshold.

Therefore, depending on a , α is defined as

$$\alpha(a) [\text{mrad}] = \begin{cases} 1.5, & \alpha < \alpha_{\min} = 1.5 \text{ mrad}, \\ a/r, & \alpha_{\min} \leq \alpha \leq \alpha_{\max}, \\ 100, & \alpha > \alpha_{\max} = 100 \text{ mrad}. \end{cases} \quad (\text{C.2})$$

Hereby, the first range corresponds to a point source, second to an intermediate extended source and the third to a large extended source.

In the particular case of a collimated beam, r is practically infinity, leading to $\alpha \leq 1.5$ mrad independent on the source size. For well collimated beams in this wavelength range, the receiver screen aperture diameter is set to $d = 50 \text{ mm}$.

The corresponding formula for AEL calculation can be found in a table provided by the norm, for the given wavelength and time base, and depending on whether or not the source is extended and whether or not the exposure time is long enough to account for the risk correction due to the eye movement. Risk correction due to the eye movement is accounted for by the viewing angle dependent time reference T_2^2 , defined as

$$T_2 [\text{s}] = \begin{cases} 10, & \alpha < \alpha_{\min}, \\ 10 \cdot 10^{(\alpha - \alpha_{\min})/98.5}, & \alpha_{\min} \leq \alpha \leq \alpha_{\max}, \\ 100, & \alpha > \alpha_{\max}. \end{cases} \quad (\text{C.3})$$

and for collimated beams $T_2 = 10 \text{ s}$. The correct formula for AEL is

$$\text{AEL} [\text{W}] = 3.9 \cdot 10^{-4} C_4 C_7, \quad (\text{C.4})$$

²In general, with increased viewing time, the retinal hazard grows. Nevertheless, it is known that then the eye movements further spread the absorbed energy, leading to different risk levels for different retinal image sizes. The break point T_2 was derived as the viewing time at which eye movements compensated for the increased theoretical risk of thermal injury for increased retinal exposure durations if the eye was immobilized. In larger extended sources, eye movements increase the retinal image more and this “break point” viewing time is longer. It is considered that for increasing angular subtense α , the breakpoint T_2 increases from 10 s for small sources to 100 s for larger sources. Beyond 100 s there is no further increase in risk of thermal injury for small and intermediate size images.

where C_4 and C_7 are wavelength dependent correction factors

$$C_4 = 10^{0.002(\lambda-700)}, \quad (\text{C.5})$$

$$C_7 = \begin{cases} 1, & 700 \text{ nm} < \lambda < 1150 \text{ nm}, \\ 10^{0.018(\lambda-1150)}, & 1150 \text{ nm} < \lambda < 1200 \text{ nm}, \\ 8, & \lambda > 1200 \text{ nm}, \end{cases} \quad (\text{C.6})$$

building up to an AEL of 0.979 mW. In principle, this value is independent on the source size, there is however a correction which directly stems from the measurement setup. As the Fig. C.2 illustrates, the detector diameter is fixed (in case of 900 nm radiation to 50 mm), regardless of a . Up to $a = 50$ mm, all radiated power falls onto the detector, whereas for larger source sizes, the source aperture stop allows only a portion of radiated power to go through the lens.

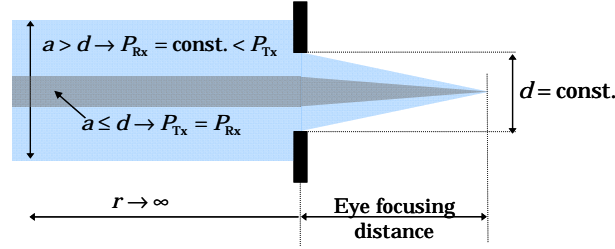


Figure C.2: Case of a collimated beam in the measurement setup.

In this case, the power falling on the detector is

$$P_{\text{Rx}} = P_{\text{Tx}} \frac{d^2}{a^2} \leq 0.979 \text{ mW}, \quad (\text{C.7})$$

so the permissible power for the source grows as

$$P_{\text{Tx}} [\text{mW}] = 0.979 \frac{a^2}{d^2}, \quad a > d \quad (\text{C.8})$$

Case 1b. $\lambda = 900$ nm, time base $t = 3 \cdot 10^4$ s, Lambertian beam

In this case, a source is not collimated and wavelength is in the near IR range, so the apparent viewing angle α depends on the apparent source size as indicated by (C.2). Measurement setup for this case considers placing the source $r = 100$ mm in front of the detector with a variable aperture stop, depending on α . The aperture of the receiver screen d should be set as

$$d(\alpha) [\text{mm}] = \begin{cases} 50, & \alpha < \alpha_{\min}, \\ 7\sqrt{\alpha_{\max}/(\alpha + 0.46)}, & \alpha_{\min} \leq \alpha \leq \alpha_{\max}, \\ 7, & \alpha > \alpha_{\max}, \end{cases} \quad (\text{C.9})$$

in order to achieve the equivalent effect as a source of size a whose distance from a 7 mm pupil is varied.

The expression for the AEL of a point source can be found in the table (in the norm) as

$$\text{AEL [W]} = \begin{cases} 7 \cdot 10^{-4} t^{-0.25} C_4 C_6 C_7, & t \leq T_2, \\ 3.9 \cdot 10^{-4} C_4 C_7, & t > T_2 \text{ and } \alpha \leq \alpha_{\min}, \\ 7 \cdot 10^{-4} T^{-0.25} C_4 C_6 C_7, & t > T_2 \text{ and } \alpha > \alpha_{\min}. \end{cases} \quad (\text{C.10})$$

The correction factors C_4 and C_7 as well as the breakpoint T_2 were introduced by (C.5), (C.6) and (C.3), respectively. In (C.10) appears another correction factor C_6 , which accounted for the risk reduction in case of extended sources

$$C_6 = \begin{cases} 1, & \alpha < \alpha_{\min}, \\ \alpha/\alpha_{\min}, & \alpha_{\min} \leq \alpha \leq \alpha_{\max}, \\ \alpha_{\max}/\alpha_{\min} = 66.7, & \alpha > \alpha_{\max}. \end{cases} \quad (\text{C.11})$$

The detected radiant power of a Lambertian source with radiation index m is determined by the active detector diameter as

$$P_{\text{Rx}} = P_{\text{Tx}} (1 - \cos^{m+1} \Phi), \quad (\text{C.12})$$

where Φ represents the half of the cross-section angle corresponding to the cone

$$\Phi = \arctan \frac{d}{2r}, \quad (\text{C.13})$$

as shown in Fig. C.3.

Calculation of the maximum permissible source power, depending on the apparent source size, results in a following discussion:

1. $a < 0.15 \text{ mm} \Rightarrow \alpha < \alpha_{\min}$
Then $d = 50 \text{ mm}$ and $\Phi = 0.245 \text{ rad}$ (C.13).
Correction factor $C_6 = 1$ and break point $T_2 = 10 \text{ s}$, $t > T_2$.
Therefore, the second case of (C.10) applies, leading to a constant AEL = 0.979 mW. The source power obtained by (C.12) is then $P_{\text{Tx}} = 16.65 \text{ mW}$.
2. $0.15 \text{ mm} \leq a < 10 \text{ mm} \Rightarrow \alpha_{\min} \leq \alpha \leq \alpha_{\max}$.
Parameters d , Φ , C_6 and T_2 vary with a , and so does the AEL and source power P_{Tx} .
3. $a = 10 \text{ mm} \Rightarrow \alpha = \alpha_{\max}$.
In this particular case, $d = 7 \text{ mm}$, $\Phi = 0.035 \text{ rad}$, $C_6 = 66.7$ and $T_2 = 100 \text{ s}$.
This leads to an AEL = 37.06 mW, and $P_{\text{Tx}} = 37 \text{ W}$.

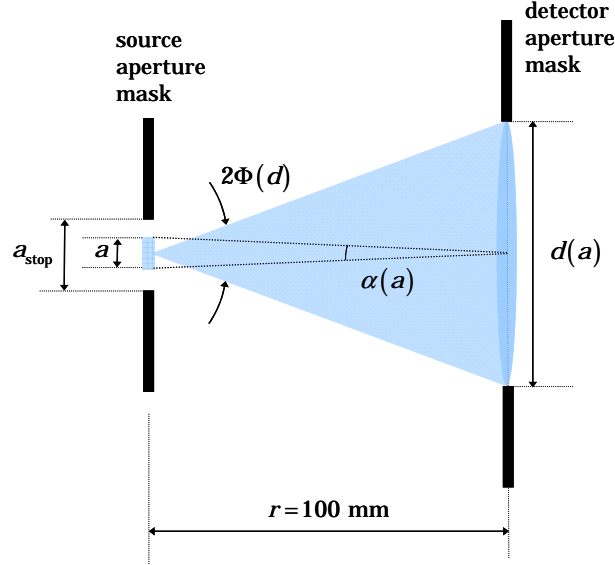


Figure C.3: Case of a Lambertian beam in the measurement setup.

4. $a > 10$ mm.

The permissible source power is now determined by

$$P_{Tx} = \text{AEL}(a = 10) \frac{a^2}{10^2}, \quad (\text{C.14})$$

due to the fact that the power passing through the source aperture remains constant. Hereby, a in [mm], and P_{Tx} and AEL in [mW] is considered. Also, $m = 1$ is assumed.

Case 2a. $\lambda = 1550$ nm, time base $t = 3 \cdot 10^4$ s, collimated beam

In this case, a detector diameter $d = 25$ mm is specified. According to the table, there is a constant AEL= 10 mW, since the risk of retinal thermal injury is independent on the source size for long wavelengths. Similarly to the case of 900 nm collimated beam (C.8), the dependence of the permissible source power on the apparent source size is

$$P_{Tx} [\text{mW}] = \begin{cases} 10, & a < d, \\ 10a^2/d^2, & a \geq d, \end{cases} \quad (\text{C.15})$$

Case 2b. $\lambda = 1550$ nm, time base $t = 3 \cdot 10^4$ s, Lambertian beam

Again, for long wavelengths, the risk of retinal thermal injury is independent on the source size, so a measurement setup with a constant detector diameter $d = 50$ mm and constant $r = 100$ mm is specified by the norm. For the point source, table

gives a unique AEL= 10 mW (as in the case 2a), and the permissible power of the Lambertian source is then given by

$$P_{\text{Tx}} [\text{mW}] = \begin{cases} 10/(1 - \cos^{m+1} \Phi), & a < a_{\text{stop}} = 10 \text{ mm}, \\ 10/(1 - \cos^{m+1} \Phi) \cdot a^2/10^2, & a \geq a_{\text{stop}}, \end{cases} \quad (\text{C.16})$$

where a constant angle Φ is determined by constant d and r (C.13).

C.2 Maximum permissible Tx power based on MPEs for skin safety

For consideration of the skin safety the norm specifies a constant detector aperture diameter $d = 3.5$ mm. According to the table, the MPE values for two considered wavelengths are

$$\text{MPE} [\text{W}/\text{m}^2] = \begin{cases} 5024, & \lambda = 900 \text{ nm}, \\ 1000, & \lambda = 1550 \text{ nm}, \end{cases} \quad (\text{C.17})$$

so the power at the detector should not exceed

$$P_{\text{Rx}} = \text{MPE} \frac{d^2 \pi}{4}, \quad (\text{C.18})$$

i.e., the permissible source power is given by

$$P_{\text{Tx}} = \begin{cases} \text{MPE} \cdot d^2 \pi / 4, & a < 3.5 \text{ mm}, \\ \text{MPE} \cdot a^2 \pi / 4, & a \geq 3.5 \text{ mm}. \end{cases} \quad (\text{C.19})$$

Appendix D

Non-Imaging Concentrator Gain

A fundamental principle of invariant etendue states that the power flow in an optical system is constant. More precisely, the product of the aperture size (a) and the numerical aperture from which the 2D system accepts light ($n \sin \phi$, n being the refraction index and 2ϕ the angular extent of the medium) is constant [73]. That means that the concentration ratio (concentrator gain) for a 2D system depicted in Fig. D.1 is

$$\frac{a'}{a''} = \frac{n'' \sin \phi''}{n' \sin \phi'}. \quad (\text{D.1})$$

Similarly, for a 3D system, the concentration ratio is

$$\left(\frac{a'}{a''} \right)^2 = \frac{n''^2 \sin^2 \phi''}{n'^2 \sin^2 \phi'}. \quad (\text{D.2})$$

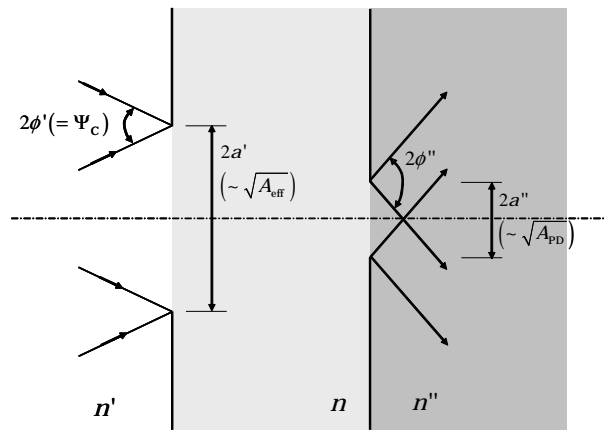


Figure D.1: The theoretical maximum concentration ratio for a 2D optical system.

These results assume that a'' is a dimension of the exit aperture large enough to permit any ray that reaches it to pass, and ϕ'' is the largest angle of all the emergent

rays. If we consider a non-imaging concentrator made of material with refraction index n placed on top of a photodetector (medium with n''), it is convenient to substitute $n'' \sin \phi''$ with $n \sin \phi$ (by Snell's refraction law), where ϕ is the angle of incidence of the ray upon the detector (inside the concentrator, ϕ is not shown in Fig. D.1). The incoming rays on the flat detector have $\phi \leq \pi/2$, so the theoretical maximum concentrator gain of a 3D non-imaging optical concentrator, is

$$g_{\max} = \frac{n^2}{\sin^2(\Psi_C/2)}, \quad (\text{D.3})$$

where the concentrator field of view $\Psi_C = 2\phi'$ (in Fig. D.1) and $n' \approx 1$ for the air are considered.

Bibliography

- [1] O. Bouchet, H. Sizun, C. Boisrobert, F. de Fornel, and P.-N. Favenneec, *Free-Space Optics: Propagation and Communication*, ISTE Ltd, 2006.
- [2] Visible Light Communications Consortium. WebSite. www.vlcc.net.
- [3] T. Komine and M. Nakagawa, “Fundamental analysis for visible-light communication system using LED lightings”, *IEEE Transactions on Consumer Electronics*, vol. 50, no. 1, pp. 100–107, 2004.
- [4] J. Grubor, S. Randel, K.-D. Langer, and J. W. Walewski, “Broadband information broadcasting using LED-based interior lighting”, *IEEE Journal of Lightwave Technology*, vol. 26, no. 24, pp. 3883–3892, December 2008.
- [5] IEEE Standard for Information technology, Telecommunications and Information exchange between systems. Local and metropolitan area networks specific requirements, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, 2007.
- [6] IEEE 802.11n Draft. Institute of Electrical and Electronics Engineers Standards Association WebSite. standards.ieee.org.
- [7] Bluetooth Special Interest Group. WebSite. www.bluetooth.com/Bluetooth/SIG.
- [8] Y. Liuqing and G. B. Giannakis, “Ultra-wideband communications: an idea whose time has come”, *IEEE Signal Processing Magazine*, vol. 21, no. 6, pp. 26 – 54, Nov. 2004.
- [9] IEEE 802.15.3a Task group. Institute of Electrical and Electronics Engineers Standards Association WebSite. standards.ieee.org.
- [10] IEEE 802.15 TG3c. Institute of Electrical and Electronics Engineers Standards Association WebSite. standards.ieee.org.
- [11] J. R. Barry and E. A. Lee, “Performance of coherent optical receivers”, *Proceedings of IEEE*, vol. 78, pp. 1369–1394, 1990.

- [12] G. Gould, S. F. Jacobs, J. T. LaTourrette, M. Newstein, and P. Rabinowitz, "Coherent detection of light scattered from a diffusely reflecting surface", *Applied Optics*, vol. 3, no. 5, pp. 648–649, May 1964.
- [13] Infrared Data Association. WebSite. www.irda.org.
- [14] D. R. Wisely, "A 1 Gb/s optical wireless tracked architecture for atm delivery", *IEE Colloquium on Optical Free Space Communication Links*, pp. 14/1–14/7, 1996.
- [15] D. R. Wisely and I. Neild, "A 100 Mb/s tracked optical wireless telepoint", *IEEE PIMRC*, pp. 694–698, 1997.
- [16] J. Bellon, M. J. N. Sibley, D. R. Wisely, and S. D. Greaves, "Hub architecture for infrared wireless networks in office environments", *IEE Proceedings Optoelectronics*, vol. 146, pp. 78–82, 1999.
- [17] V. Jungnickel, T. Haustein, A. Forck, and C. von Helmolt, "155 Mb/s wireless transmission with imaging infrared receiver", *Electronics Letters*, vol. 37, pp. 314–315, 2001.
- [18] IL 60201 Spectrix Corp., Evanston.
- [19] Institute of Electrical and Electronics Engineers. Standards Association WebSite. standards.ieee.org.
- [20] M. D. Audeh, J. M. Kahn, and J. R. Barry, "Performance of pulse position modulation on measured non-directed indoor infrared channels", *IEEE Transactions on Communications*, vol. 4, no. 6, pp. 654–659, 1996.
- [21] H. Park and J. R. Barry, "Performance of multiple pulse position modulation on multipath channels", *IEE Proceedings on Optoelectronics*, vol. 143, no. 6, pp. 360–364, 1996.
- [22] D.-S. Shiu and J. M. Kahn, "Differential pulse-position modulation for power-efficient optical communication", *IEEE Transactions on Communications*, vol. 47, no. 8, pp. 1201–1210, 1999.
- [23] Z. Ghassemlooy and A. R. Hayes, "Digital pulse interval modulation for IR communication systems - a review", *International Journal of Communication Systems*, vol. 13, pp. 519 – 536, 2000.
- [24] N. M. Aldibbiat, Z. Ghassemlooy, and R. McLaughlin, "Dual header pulse interval modulation for dispersive indoor optical wireless communication systems", *IEE Proceedings - Circuits, Devices and Systems*, vol. 149, no. 3, pp. 187–192, 2002.

- [25] J. B. Carruthers and J. M. Kahn, "Angle diversity for nondirected wireless infrared communication", *IEEE Transactions on Communications*, vol. 48, no. 6, June 2000.
- [26] G. W. Marsh and J. M. Kahn, "Performance evaluation of experimental 50 Mb/s diffuse infrared wireless link using on-off keying with decision-feedback equalization", *IEEE Transactions on Communication*, vol. 44, no. 11, pp. 1496–1504, 1996.
- [27] K. K. Wong, T. O'Farrell, and M. Kiatweerasakul, "Infrared wireless communication using spread spectrum techniques", *IEE Proceedings Optoelectronics*, vol. 147, no. 4, pp. 308–314, 2000.
- [28] K. K. Wong and T. O'Farrell, "Spread spectrum techniques for indoor wireless IR communications", *IEEE Wireless Communications*, vol. 10, no. 2, pp. 54–63, 2003.
- [29] S. T. Jivkova and M. Kavehrad, "Multispot diffusing configuration for wireless infrared access", *IEEE Transactions on Communications*, vol. 48, pp. 970–978, June 2000.
- [30] M. Castillo-Vazquez and A. Puerta-Notario, "Self-orienting receiver for indoor wireless infrared links at high bit rates", *IEEE VTC Spring*, vol. 3, pp. 1600–1604, 2003.
- [31] V. Jungnickel, A. Forck, T. Haustein, U. Krueger, V. Pohl, and C. von Helmolt, "Electronic tracking for wireless infrared communication", *IEEE Transactions on Wireless Communications*, vol. 2, 2003.
- [32] H. Hashemi, G. Yun, M. Kavehrad, F. Behbahani, and P. A. Galko, "Indoor propagation measurements at infrared frequencies for wireless local area networks applications", *IEEE Transactions on Vehicular Technology*, vol. 43, pp. 562–575, 1994.
- [33] S. Nönning, "Kanalmessungen zur Infrarotübertragung im Indoor Bereich", Diplomarbeit, Technische Universität Berlin, 2000.
- [34] M. Matsuo, T. Ohtsuki, and I. Sasase, "Rate-adaptive indoor infrared wireless communication systems using punctured convolutional codes and adaptive PPM", in *IEEE PIMRC*, 1998.
- [35] L. Diana and J. M. Kahn, "Rate-adaptive modulation techniques for infrared wireless communications", in *IEEE ICC*, 1999.
- [36] T. Ohtsuki, "Rate-adaptive indoor infrared wireless communication systems using repeated and punctured convolutional codes", *IEEE Communications Letters*, vol. 4, no. 2, February 2000, Serial transmission, adaptive coding.

- [37] A. Garcia-Zambrana and A. Puerta-Notario, "Rate-adaptive indoor wireless links using OOK formats with alternate-position gaussian pulses", *Electronics Letters*, vol. 37, no. 23, pp. 1409–1411, November 2001.
- [38] J. M. Garrido-Balsells, A. Garcia-Zambrana, and A. Puerta-Notario, "Performance evaluation of rate-adaptive transmission techniques for optical wireless communications", in *IEEE VTC Spring*, 2004, pp. 914–918.
- [39] K. Akhavan, M. Kavehrad, and S. Jivkova, "High-speed power-efficient indoor wireless infrared communication using code combining - part I and II", *IEEE Transactions on Communications*, vol. 50, no. 7, July 2002.
- [40] M. Castillo-Vazquez, A. Garcia-Zambrana, and A. Puerta-Notario, "Angle diversity with rate-adaptive transmission using repetition coding and variable silence periods for wireless infrared communications", in *IEEE VTC Spring*, 2004, vol. 1, pp. 415–419.
- [41] J. R. Barry, *Wireless Infrared Communication*, Kluwer Academic Publishers, 1994.
- [42] J. B. Carruthers and J. M. Kahn, "Multiple-subcarrier modulation for nondirected wireless infrared communications", *IEEE Journal of Selected Areas in Communications*, vol. 14, no. 3, 1996.
- [43] T. Ohtsuki, "Multiple-subcarrier modulation in optical wireless communications", *IEEE Communications Magazine*, March 2003.
- [44] S. Hara and R. Prasad, *Multicarrier Techniques for 4G Mobile Communications*, Artech House, 2003.
- [45] ITU-T Recommendation G.992.1 (G.DMT), 1999.
- [46] J. M. Cioffi, "A multicarrier primer", *ANSI Contribution T1E1*, pp. 4/91–157, November 1991.
- [47] L. Hanzo, T. Münster, T. Keller, and B. J. Choi, *OFDM and MC-CDMA for Broadband Multi-user Communications WLANs and Broadcasting*, John Wiley and Sons, 2003.
- [48] F. Xiong, *Digital Modulation Techniques*, Artech House, second edition, 2006.
- [49] R. You and J. M. Kahn, "Upper-bounding the capacity of optical IM/DD channels with multiple-subcarrier modulation and fixed bias using trigonometric moment space method", *IEEE Transactions on Information Theory*, vol. 48, no. 2, pp. 514–523, February 2002.

- [50] S. Haykin, *Communications Systems*, John Wiley & Sons, Inc., 1994.
- [51] J. M. Kahn and J. R. Barry, “Wireless infrared communications”, *Proceedings of IEEE*, vol. 85, pp. 265–298, 1997.
- [52] W. C. Jakes, *Microwave Mobile Communications*, IEEE Press, 1994.
- [53] J. G. Proakis, *Digital Communications*, McGraw-Hill, 2001.
- [54] H. Beyer, “Untersuchung des innenraum-Infrarot-Übertragungskkanals durch Anwendung des Ray-Tracing”, Diplomarbeit, Fachhochschule für Technik und Wirtschaft Berlin, 2007.
- [55] V. Jungnickel, V. Pohl, S. Noenning, and C. von Helmolt, “A physical model of the wireless infrared communication channel”, *IEEE Journal of Selected Areas in Communications*, vol. 20, 2002.
- [56] A. V. Oppenheim and R. W. Schaffer, *Discrete-Time Signal Processing*, Prentice-Hall Signal Processing Series, 1989.
- [57] V. Pohl, V. Jungnickel, and C. von Helmolt, “A channel model for wireless infrared communications”, *IEEE PIMRC*, 2000.
- [58] J. B. Carruthers and J. M. Kahn, “Modeling of nondirected wireless infrared channels”, *IEEE Transactions on Communications*, vol. 45, pp. 1260–1268, October 1997.
- [59] M. Abtahi and H. Hashemi, “Simulation of indoor propagation channel at infrared frequencies in furnished office environments”, *IEEE PIMRC*, vol. 6, pp. 306–310, 1995.
- [60] F. R. Gfeller and U. Bapst, “Wireless in-house data communication via diffuse infrared radiation”, *Proceedings of the IEEE*, vol. 67, pp. 1474–1486, 1979.
- [61] V. Pohl, V. Jungnickel, and C. von Helmolt, “Integrating-sphere diffuser for wireless infrared communication”, *IEEE Proceedings in Optoelectronics*, vol. 147, 2000.
- [62] A. J. C. Moreira, R. T. Valadas, and A. M. de Oliveira Duarte, “Optical interference produced by artificial light”, *Wireless Networks*, vol. 3, pp. 131–140, 1997.
- [63] H.-A. Bachor and P. T. H. Fisk, “Quantum noise - a limit in photodetection”, *Applied Physics B*, vol. 49, pp. 291–300, 1989.
- [64] G. P. Agrawal, *Fiber-Optic Communication Systems*, John Wiley and Sons, Inc., 1997.

- [65] Z. Wang and G. B. Giannakis, “Wireless multicarrier communications: Where fourier meets shannon”, *IEEE Signal Processing Magazine*, vol. 17, no. 3, pp. 29–48, May 2000.
- [66] A. Goldsmith, *Wireless Communications*, Cambridge University Press, 2005.
- [67] P. L. Eardkey, D. R. Wisley, D. Wood, and P. McKee, “Holograms for optical wireless LANs”, *IEE Proceedings on Optoelectronics*, vol. 143, pp. 365–369, 1996.
- [68] “International standard IEC 60825-1: Safety of laser products”, Tech. Rep., International Electrotechnical Commission, Edition 08 / 2001.
- [69] “International standard IEC 60825-1: Safety of laser products”, Tech. Rep., International Electrotechnical Commission, Edition 2007.
- [70] K. Petermann, *Laser Diode Modulation and Noise*, Kluwer Academic Publishers, 1988.
- [71] N. R. Draper and H. Smith, *Applied Regression Analysis*, Wiley Series in Probability and Statistics, 1998.
- [72] J. R. Barry and J. M. Kahn, “Link design for nondirected wireless infrared communications”, *Applied Optics*, vol. 34, no. 19, pp. 3764–3776, July 1995.
- [73] W. T. Weltford and R. Winston, *High-Collection Non-Imaging Optics*, Academic Press, 1989.
- [74] M. Wolf, *Optischer Empfänger mit geringer spektraler Breite für den Einsatz in der leitungsungebundenen optischen Übertragung*, Deutsches Patentamt (DE 10105359.2), 2001.
- [75] E. Voges and K. Petermann, *Optische Kommunikationstechnik*, Springer, 2002.
- [76] M. Wolf, *Zur breitbandigen Infrarot-Indoorkommunikation*, PhD thesis, Technischen Universität Illmenau, 2002.
- [77] K. Petermann, “Vorlesungskript: Einführung in die optische nachrichtentechnik”, 2003.
- [78] R. G. Smith and S. D. Personick, “Receiver design for optical fiber communication systems”, in *Semiconductor Deices for Optical Communication*, H. Kressel, Ed. Springer Verlag, 1980.
- [79] Gerhard Grau, *Optische Nachrichtentechnik*, Springer-Verlag, 1981.

- [80] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., 1991.
- [81] C. E. Shannon, “A mathematical theory of communications”, *The Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, July, October 1948.
- [82] S. Haykin, *Adaptive Filter Theory, Appendix C*, Prentice Hall, 2002.
- [83] A. A. Farid and S. Hranilovic, “Upper and lower bounds on the capacity of wireless optical intensity channels”, *IEEE International Symposium on Information Theory*, 2007.
- [84] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [85] W. Rudin, *Principles of Mathematical Analysis*, Mc-Graw Hill, 1976.
- [86] I. J. Schoeneberg, “An isoperimetric inequality for closed curves convex in even-dimensional euclidean spaces”, *Acta Mathematica*, vol. 91, pp. 143–164, 1954.
- [87] S. R. Lay, *Convex Sets and Their Applications*, Pure & Applied Mathematics, A Wiley-Interscience Series, 1982.
- [88] H. S. M. Coxeter, *Regular Polytopes*, The Macmillan Company, New York, 1963.
- [89] K. Fan, “On positive definite sequences”, *Annals of Mathematics*, vol. 47, no. 3, pp. 593–607, July 1946.
- [90] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*, SIAM: Society for Industrial and Applied Mathematics, 2001.
- [91] R. Gaudino, G. Bosco, A. Bluschke, O. Hofmann, N. Kiss, M. Matthews, P. Rietzsch, S. Randel, J. Lee, and F. Breyer, “On the ultimate capacity of SI-POF links and the use of OFDM: Recent results from the POF-ALL project”, *ICPOF*, 2007.
- [92] R. F. H. Fischer and J. B. Huber, “A new loading algorithm for discrete multitone transmission”, in *IEEE GLOBECOM*, Nov 1996, pp. 724–728.
- [93] O. Gonzalez, S. Rodriguez, R. Perez-Jimenez, F. Delgado, and A. Ayala, “Multi-user adaptive orthogonal frequency-division multiplexing system for indoor wireless optical communications”, *IET Optoelectronics*, vol. 1, no. 2, pp. 68 – 76, April 2007.
- [94] Hughes-Hartogs, “Ensemble modem structure for imperfect transmission media”, *United States Patent Nr. 4,731,816*, 1988.

- [95] P. S. Chow, J. M. Cioffi, and J. A. C. Bingham, "A practical discrete multitone transceiver loading algorithm for data transmission over spectrally shaped channels", *IEEE Transactions on Communications*, vol. 43, pp. 773–775, Feb/Mar/Apr 1995.
- [96] B. S. Krongold, K. Ramchandran, and D. L. Jones, "Computationally efficient optimal power allocation algorithms for multicarrier communication systems", *IEEE Transactions on Communications*, vol. 48, no. 1, pp. 23–27, 2000.
- [97] R. You and J. M. Kahn, "Average power reduction techniques for multiple-subcarrier intensity-modulated optical signals", *IEEE Transactions on Communications*, vol. 49, no. 12, pp. 2164–2171, December 2001.
- [98] H. Yamaguchi, T. Ohtsuki, and I. Sasase, "Multiple subcarrier modulation for infrared wireless systems using punctured convolutional codes and variable amplitude blocks", *IEICE Transactions on Communications*, vol. E86-B, no. 8, August 2003.
- [99] S. Teramoto and T. Ohtsuki, "Multiple-subcarrier optical communication systems with subcarrier signal-point sequence", *IEEE Transactions on Communications*, vol. 53, no. 10, pp. 1738–1743, October 2005.
- [100] N. Kitamoto and T. Ohtsuki, "Parallel combinatory multiple-subcarrier optical communication systems", *IEEE PIMRC*, 2002.
- [101] S. Teramoto and T. Ohtsuki, "Multiple-subcarrier optical communication system with peak reduction carriers", *IEICE Transactions on Communication*, vol. E87-B, no. 11, November 2004.
- [102] X. Li, R. Mardling, and J. Armstrong, "Channel capacity of IM/DD optical communications systems and of ACO-OFDM", *IEEE ICC*, 2007.
- [103] V. Pohl, V. Jungnickel, and C. von Helmolt, "Optimal length of the training and data phase in continuous flat fading MIMO channels", *IEEE International Conference on Communications*, pp. 3153–3157, 2004.
- [104] R. M. Gray, *Toeplitz and Circulant Matrices: A Review*, Now Publishers Inc, 2006.
- [105] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, 1991.

Publications of the Author

Book contributions

1. D. O' Brien et al, "Visible Light Communications", to appear in "Short-Range Wireless Communications" by R. Krämer and Marcos D. Katz (Eds.), John Wiley and Sons Ltd., 2009

Journal papers

1. J. Grubor, S. Randel, K.-D. Langer and J. W. Walewski, "Broadband information broadcasting using LED-based interior lighting", *IEEE Journal of Lightwave Technology*, Vol. 26, pp. 3883-3892, December 2008.
2. X. Li, J. Grubor, V. Jungnickel and J. Armstrong, "Channel Capacity of ACO-OFDM in Optical Wireless Applications", accepted for publication in *IEEE Transactions on Communications*.
3. J. Grubor and K.-D. Langer, "Efficient Signal Processing in OFDM-based Indoor Optical Wireless Links", to be published in *Journal of Networks, Academy Publisher*, February 2010.
4. J. Vucic, C. Kottke, S. Nerreter, A. Buettner, K.-D. Langer, and J. W. Walewski "White-Light Wireless Transmission at 200+ Mbit/s Net Data Rate by Use of Discrete-Multitone Modulation", accepted for publication in *IEEE Photonics Technology Letters*.

Invited papers

1. K.-D. Langer, J. Grubor and K. Habel, "Promising evolution paths for optical access networks", *IEEE International Conference on Transparent Optical Networks (ICTON)*, Vol. 1, pp. 202-207, 2004, (invited).
2. K.-D. Langer and J. Grubor, "Recent developments in optical wireless communications using infrared and visible light", *IEEE ICTON*, Vol. 3, pp.146-151, Paper Tu.D3.1, 2007 (invited).
3. J. Grubor, V. Jungnickel and K.-D. Langer, "Adaptive optical wireless OFDM system with controlled asymmetric clipping", *IEEE Asilomar Conference on Signals, Systems and Computers*, 2007 (invited).
4. K.-D. Langer, J. Grubor, O. Bouchet, M. El Tabach, J.W. Walewski, S. Randel, M. Franke, S. Nerreter, D.C. O'Brien, G.E. Faulkner, I. Neokosmidis, G. Ntogari and M. Wolf, "Optical wireless communications for broadband access in home area networks". *IEEE Int. Conference on Transparent Optical Networks (ICTON)*, vol. 4, pp. 149-154, We.B4.1, 2008 (invited).

White papers

1. H. Le-Minh, D.C. O'Brien, L. Zeng, O. Bouchet, S. Randel, J. Walewski, J.A.R. Borges, K.-D. Langer, J. Grubor, K. Lee and E. T. Won, "Short-range Visible Light Communications", *White-Paper in Wireless World Research Forum (WWRF) 19*, Chennai, India, Nov. 2007.
2. D.C. O'Brien, L. Zeng, H. Le-Minh, G. Faulkner, O. Bouchet, S. Randel, J. Walewski, J.A. R. Borges, K.-D. Langer, J. Grubor, K. Lee and E. T. Won, "Visible Light Communications", *White-Paper in Wireless World Research Forum (WWRF) 20*, Ottawa, Canada, Apr. 2008.

Conference papers

1. J. Vathke, M. Schlosser, J. Grubor and K.-D. Langer, "CWDM in breitbandigen Zugangsnetzen", *ITG Fachbericht 180, Kommunikationskabelnetze*, pp. 81-90, 2003.
2. J. Grubor, M. Schlosser and K.-D. Langer, "Feasibility of protected rings in optical access networks", *ITG Fachbericht 182, Photonische Netze*, pp. 155-162, 2004.
3. J. Grubor, M. Schlosser and K.-D. Langer, "Protected ring network for future optical access domain", *European Conference on Networks and Optical Communications (NOC)*, pp. 106-113, 2004.
4. J. Grubor, V. Jungnickel and K.-D. Langer, "Capacity analysis in indoor wireless infrared communication using adaptive multiple subcarrier transmission", *IEEE International Conference on Transparent Optical Networks (ICTON)*, Vol. 2, pp. 171-174, 2005.
5. J. Grubor, V. Jungnickel and K.-D. Langer "Rate-adaptive multiple subcarrier based transmission for broadband infrared wireless communication", *National Fiber Optic Communication Conference (NFOEC)*, paper NThG2, 2006.
6. J. Grubor, V. Jungnickel and K.-D. Langer, "Adaptive-modulation technique in wireless infrared indoor communications", *ITG Fachbericht 193, Photonische Netze*, pp. 193-200, 2006.
7. J. Grubor, V. Jungnickel and K.-D. Langer, "Bit-loading for modulation-adaptive transmission in infrared wireless indoor communication", *IET International Conference on Access Technologies*, pp. 9-12, 2006.
8. J. Grubor, O.C. Gaete Jamett, J.W. Walewski, S. Randel and K.-D. Langer, "High-speed wireless indoor communication via visible light", *ITG Fachbericht 198, Breitbandversorgung in Deutschland - Wie schaffen Wir den Anschluss?*, pp. 203-208, 2007.

9. D. O'Brien, O. Bouchet, S. Randel, J. Walewski, K.-D. Langer, J. Grubor, J.A. Rabadan Borges, R. Perez-Jimenez, E.T. Won, "Briefing on Visible-Light Communications: Update and Key Questions for Whitepaper", *Wireless World Research Forum (WWRF)* 18, June 2007, Espoo, Finland.
10. J. Grubor, S.C.J. Lee, K.-D. Langer, T. Koonen and J.W. Walewski, "Wireless high-speed data transmission with phosphorescent white-light LEDs", *European Conference and Exhibition on Optical Communication (ECOC)*, Vol. 6, Post-Deadline Paper PD3.6, 2007.
11. J. Grubor, S. Randel, K.-D. Langer and J.W. Walewski, "Bandwidth-efficient indoor optical wireless communications with white light-emitting diodes", *IEEE/ IET Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*, pp. 165-169, 2008.
12. O. Bouchet, M. El Tabach, M. Wolf, D.C. O'Brien, G.E. Faulkner, J.W. Walewski, S. Randel, M. Franke, S. Nerreter, K.-D. Langer, J. Grubor and T. Kamalakis, "Hybrid Wireless Optics (HWO): Building the next-generation home network", *IEEE/ IET Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*, pp. 283-287, 2008.
13. D.C. O'Brien, G. Faulkner, H. Le-Minh, O. Bouchet, M.E. Tabach, M. Wolf, J. W. Walewski, S. Randel, S. Nerreter, M. Franke, K.-D. Langer, J. Grubor and T. Kamalakis, "Home access networks using optical wireless transmission", *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Cannes, France, Sept. 2008.
14. K.-D. Langer, K. Habel, J. Grubor and A. Koher, "Hochratige Datenübertragung im Heimnetzbereich", *ITG Fachbericht 210, Breitbandversorgung in Deutschland - mehr Bandbreite für alle?*, pp. 199-206, 2008.
15. K.-D. Langer, J. Vucic, C. Kottke, L. Fernández del Rosal S. Nerreter, J.W. Walewski, "Advances and Prospects in High-Speed Information Broadcast using Phosphorescent White-Light LEDs", *IEEE ICTON*, Mo.B5.3, 2009.
16. A. Paraskevopoulos, J. Vucic, S.-H. Voß, K.-D Langer, "Optical Free-Space Communication Systems in the Mbps to Gbps Range, Suitable for Industrial Applications" accepted for *IEEE ISOT* 2009.
17. J. Vucic, C. Kottke, S. Nerreter, K. Habel, A. Buettner, K.-D. Langer and J. W. Walewski, "125 Mbit/s over 5 m Wireless Distance by Use of OOK-Modulated Phosphorescent White LEDs", accepted for *ECOC* 2009.

Oral presentations

1. C. Caspar, G. Großkopf, J. Grubor, K. Habel, K.-D. Langer, D. Rohde, M. Schlosser, M. Seimetz and J. Vathke, “Drahtlos und breitbandig vom Schreibtisch ins Glasfaser-Festnetz”, *ITG Workshop Zukunft der Netze*, Kaiserslautern, Deutschland, 2004.
2. J. Grubor, “Using LEDs twice - for illumination and information broadcasting”, presented at the *OPTECBB Summer School*, Berlin, Germany, 2008.
3. J. W. Walewski, K.-D. Langer and J. Vucic, “The OMEGA use case”, presented at IEEE P802.15 Working Group for Wireless Personal Area Networks (WPANs), Vancouver, Canada, March, 2009.

Patents of the Author

1. J. Grubor, V. Jungnickel, K.-D. Langer, C. v. Helmolt, “Dynamisches datenratenadaptives Signalverarbeitungsverfahren in einem drahtlosen Infrarot-Datenübertragungssystem”, PCT/DE2006/000750, 2006.

Biography

Jelena Vučić (maiden name Grubor) was born in Belgrade, Serbia, in 1979. She received the Diploma in Electrical Engineering with emphasis on Telecommunications, at Belgrade University in 2002. In 2003, she joined the Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institut, in Berlin, Germany, where she works as a research assistant in the Department of Photonic Networks and Systems. Jelena finished her Ph.D. thesis in 2009 at Technical University of Berlin. The main topic of her thesis was the investigation of advanced signal processing techniques in broadband optical wireless systems. Her current research interests include optical access and indoor networks and optical wireless systems.