

ITG-Fachbericht **305**

# Photonische Netze

Beiträge der 23. ITG-Fachtagung  
18. – 19. Mai 2022, Berlin



**Universität Stuttgart**

# Photonische Netze

Beiträge der 23. ITG-Fachtagung  
18. – 19. Mai 2022 in Berlin

Veranstalter:

Informationstechnische Gesellschaft im VDE (ITG)  
Fachausschuss KT3 „Optische Nachrichtentechnik“ und  
ITG-Fachgruppe KT3.3 „Photonische Netze“  
mit freundlicher Unterstützung der Universität Stuttgart

Wissenschaftliche Tagungsleitung:

Prof. Dr.-Ing. Andreas Kirstädter,  
Institut für Kommunikationsnetze und Rechnersysteme,  
Universität Stuttgart

### **Bibliografische Information der Deutschen Nationalbibliothek**

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.dnb.de> abrufbar.

ISBN 978-3-8007-5877-7 (CD-ROM)

ISBN 978-3-8007-5878-4 (E-Book)

ISSN 0932-6022

© 2022 VDE VERLAG GMBH · Berlin · Offenbach, Bismarckstraße 33, 10625 Berlin

[www.vde-verlag.de](http://www.vde-verlag.de)

Alle Rechte vorbehalten

Das Werk ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlags unzulässig und strafbar. Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbeschreibungen etc. berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Markenschutz-Gesetzgebung als frei zu betrachten wären und von jedermann benutzt werden dürfen. Aus der Veröffentlichung kann nicht geschlossen werden, dass die beschriebenen Lösungen frei von gewerblichen Schutzrechten (z. B. Patente, Gebrauchsmuster) sind. Eine Haftung des Verlags für die Richtigkeit und Brauchbarkeit der veröffentlichten Programme, Schaltungen und sonstigen Anordnungen oder Anleitungen sowie für die Richtigkeit des technischen Inhalts des Werks ist ausgeschlossen. Die gesetzlichen und behördlichen Vorschriften sowie die technischen Regeln (z. B. das VDE-Vorschriftenwerk) in ihren jeweils geltenden Fassungen sind unbedingt zu beachten.

CD-Produktion: DMS – Disk Media Service, Berlin

Produced in Germany

## Mitglieder des ITG-Fachausschusses KT3

### Optische Nachrichtentechnik

Prof. Dr.-Ing. Thomas Bauschert	Technische Universität Chemnitz
Dipl.-Ing. Matthias Berger	Finisar Corporation
Dr.-Ing. Dirk Breuer	Deutsche Telekom AG
Prof. Dr.-Ing. Christian-Alexander Bunge	Hochschule der Deutschen Telekom
Dr.-Ing. Jörg-Peter Elbers	ADVA Optical Networking SE
Prof. Dr.-Ing. habil. Ulrich Fischer-Hirchert	Hochschule Harz
Dr. Chris Fludger	Infinera GmbH
Prof. Dr.-Ing. Ronald Freund	Technische Universität Berlin, Fraunhofer Heinrich-Hertz-Institut
Dr.-Ing. Dirk Giggenbach	Dt. Zentrum für Luft- und Raumfahrt e.V.
Dr.-Ing. Andreas Gladisch	Deutsche Telekom AG
Prof. Dr.-Ing. Hans-Joachim Grallert	Fraunhofer Heinrich-Hertz-Institut
Dr.-Ing. Helmut Grießer	ADVA Optical Networking SE
Dr.-Ing. Gernot Göger	Huawei Technologies
Prof. Dr.-Ing. Norbert Hanik	Technische Universität München
Prof. Dr. techn. Admela Jukan	Technische Universität Braunschweig
Prof. Dr.-Ing. Andreas Kirstädter	Universität Stuttgart
Prof. Dr.-Ing. Peter Krummrich	Technische Universität Dortmund
Prof. Dr. sc. nat. Jürg Leuthold	ETH Zürich
Dipl.-Ing. Christian Neumeyr	VERTILAS GmbH
Prof. Dr.-Ing. Reinhold Noé	Universität Paderborn
Prof. Dr.-Ing. Klaus Petermann	Technische Universität Berlin
Prof. Dr.-Ing. Stephan Pachnicke	Christian-Albrechts-Universität zu Kiel
Prof. Dr.-Ing. Sebastian Randel	Karlsruher Institut für Technologie
Prof. Dr.-Ing. Werner Rosenkranz	Christian-Albrechts-Universität zu Kiel
Prof. Dr.-Ing. Christian Schäffer	Helmut Schmidt Universität
Prof. Dr.-Ing. Christoph Scheytt	Universität Paderborn
Prof. Dr.-Ing. Bernhard Schmauss	Friedrich-Alexander-Universität Erlangen-Nürnberg
Dr.-Ing. Christoph Schulien	Ranovus GmbH
Dipl.-Ing. Bernd Sommerkorn-Krombholz	Infinera GmbH
Prof. Dr.-Ing. Andreas Stöhr	Universität Duisburg-Essen
Dipl.-Phys. Andreas Umbach	AUCCEPT Consulting GmbH
Dr. Thomas Weidlich	Seim & Partner
Prof. Dr.-Ing. Lars Zimmermann	Technische Universität Berlin

# Inhaltsverzeichnis

## Keynotes

Andreas Kirstädter

- 01 Roadmap and standards of optical interconnect and module technologies for next generation networks**  
Ross Saunders (ADVA Optical Networking, München)
- 02 Test and Validation in the Emerging World of Disaggregation**  
Andreas Gladisch (Deutsche Telekom AG)
- 03 6G als Schlüsseltechnologie für eine hochvernetzte Gesellschaft**  
Hans Schotten (DFKI, Kaiserslautern)
- 04 Optical Communication in „New Space“ – Constellations, Networks, Synergies**  
Mark Gregory (TESAT, Backnang)

## Session: Networks

- 05 Protected Connection Provisioning with Low Availability Overfulfillment in Meshed Core Networks ..... 6**  
Tobias Enderle (Universität Stuttgart)
- 06 Towards a Hybrid Architecture by Introducing Coherent Pluggable Transceivers in IP-Optical Core Networks with Optical Cross-Connects..... 14**  
Filippos Christou, Tobias Enderle, Arthur Witt (Universität Stuttgart, )
- 07 Control-Plane Traffic Modelling for Connection Management in T-SDN Optical Networks using Transport PCE and OpenROADM ..... 22**  
Shabnam Sultana\*, Ronald Romero Reyes, Thomas Bauschert (Technische Universität Chemnitz; \*und Highstreet Technologies GmbH, Berlin)

## Session: Systems & Components

- 08 Optical Satellite Technologies for 6G (Invited)**  
Nicolas Perlot (Fraunhofer Heinrich-Hertz-Institut, Berlin)
- 09 Verschränkungs-basierte Quantenkommunikation in Freistrahl- und Fasernetzen (Invited)**  
Fabian Steinlechner (Fraunhofer IOF, Jena)
- 10 Design and performance characteristics of extended C-band amplifiers..... 31**  
Lutz Rapp (ADVA Optical Networking,Meiningen)

<b>11</b>	<b>Energy Efficiency Limits to ICT Equipment Lifetime.....</b>	<b>38</b>
	Klaus Grobe (ADVA Optical Networking, München)	
<b>12</b>	<b>Two-Sided Endless Optical Polarization Control with 2 x 100 krad/s Tracking Speed.....</b>	<b>46</b>
	Reinhold Noé, Benjamin Koch (Universität Paderborn; Novoptel GmbH, Paderborn)	

### Session: Transmission I

<b>13</b>	<b>On Symmetry Properties of Transmission Matrices for Concatenations of Multi Mode Fiber Sections with Linear Coupling.....</b>	<b>49</b>
	Christian M. Spenner, Klaus Petermann*, Peter M. Krummrich (Technische Universität Dortmund *Technische Universität Berlin)	
<b>14</b>	<b>Introducing <math>\gamma</math>-lifting for Learning Nonlinear Pulse Shaping in Coherent Optical Communication.....</b>	<b>57</b>
	Tim Uhlemann, Alexander Span, Sebastian Dörner, Stephan ten Brink (Universität Stuttgart)	
<b>15</b>	<b>Time Adaptive Probabilistic Shaping for Combined Optical/THz Links.....</b>	<b>65</b>
	In-Ho Baek <sup>1</sup> , Frederik Bart <sup>2</sup> , Robert Elschner <sup>1</sup> , Fred Meier <sup>2</sup> , David Hellmann <sup>2</sup> , Andreas Maaßen <sup>1</sup> , Colja Schubert <sup>1</sup> , Ronald Freund <sup>1</sup> ( <sup>1</sup> Fraunhofer HHI, Berlin; <sup>2</sup> Technische Universität Berlin)	
<b>16</b>	<b>Experimental Investigation of Information Bit Scrambling for Physical-Layer Security in Coherent Fiber-Optic Systems.....</b>	<b>73</b>
	Carsten Schmidt-Langhorst, Johannes Pfeiffer, Robert Elschner, Robert Emmerich, Fabian Chowanek, In-Ho Baek, Robert F.H. Fischer, Colja Schubert (Fraunhofer Heinrich-Hertz-Institut, Berlin; Universität Ulm)	

### Session: Transmission II

<b>17</b>	<b>Experimental validation of exact component parameter agnostic QoT estimation using spectral data-driven LSTM over a 265.2 km SSMF link.....</b>	<b>79</b>
	Lars Kruse, Stephan Pachnicke (Christian-Albrechts-Universität, Kiel)	
<b>18</b>	<b>High Bandwidth and Ultra Low-Latency Near IR Communication Network for CMOS-compatible Integrated Photonics Chips.....</b>	<b>84</b>
	Javier Acevedo, Shahryar Sabouri, Shiwei Shen, Christoph Keller, Joerg Hopfe*, Stefan Reichmuth, Patrick Hobi*, Marco Dietrich*, Kambiz Jamshidi, Frank H.P. Fitzek (Technische Universität Dresden; *albis-elcon system Germany GmbH, Hartmannsdorf)	
<b>19</b>	<b>Distributed Sensing of Single Mode Fibers with Correlation Techniques.....</b>	<b>92</b>
	Florian Azendorf, André Sandmann, Michael Eiselt, Bernhard Schmauss* (ADVA Optical Networking, Meiningen; *Friedrich-Alexander-Universität Erlangen-Nürnberg)	

# Protected Connection Provisioning with Low Availability Overfulfillment in Meshed Core Networks

Tobias Enderle

*Institute of Communication Networks and Computer Engineering (IKR)*

*University of Stuttgart*

Stuttgart, Germany

tobias.enderle@ikr.uni-stuttgart.de

**Abstract**—Network operators sell connection services to their customers. The corresponding service contracts typically include a service level agreement (SLA) which defines a guaranteed availability level per month. An availability exceeding the SLA level means that the network operator complies with the SLA. A lower availability results in contractual penalties. Therefore, the probability of SLA compliance is a key figure for the operator. To adjust the compliance probability, many operators apply protection mechanisms, which tie up precious network resources to provide backup capacity. Therefore, it is desirable to dedicate only as many network resources to protection as necessary to reach a sufficiently high probability of SLA compliance—the operator’s self-defined compliance target. However, in practice, this is difficult to realize because using protection, the amount of resources and, consequently, the compliance probability cannot be selected continuously. Adding protection to a connection service makes its compliance probability jump up, possibly to a level far above the operator’s compliance target. The result is overfulfillment at the cost of precious network resources. In this paper, we propose an admission and routing approach that reduces said overfulfillment, frees network resources and by that allows more services to be accommodated in the network. We use a stochastic approach to estimate a service’s probability of SLA compliance. Probability that exceeds the operator’s compliance target is accumulated as surplus and allows other services to be accepted with a compliance probability below the operator’s original compliance target. With this approach, the resulting SLA compliance ratio over all services matches the compliance target closely, i.e., the overfulfillment is reduced. We evaluate our mechanism in a simulation study covering several core network topologies. It is shown that the availability overfulfillment can be reduced or even eliminated and that the service blocking ratio can be decreased significantly.

**Index Terms**—Availability management, Compliance probability, Interval availability, Overfulfillment, Routing, Service level agreement

## I. INTRODUCTION

The goal of every network provider is to achieve sustainable profit. In the past, the sale of connection services generated enough revenue. However, traffic demand is ever increasing and network operators have to expand their networks constantly

This work has been performed in the framework of the CELTIC-NEXT EUREKA project AI-NET-ANTILLAS (Project ID C2019/3-3), and it is partly funded by the German BMBF (Project ID 16KIS1312). The authors alone are responsible for the content of the paper.

to provide enough capacity. In recent years, though, network expansion has become increasingly costly, due to disparities between the growth rate of traffic demand and that of transmission technologies [1]. Additionally, network services are more and more considered a commodity and prices are falling. Therefore, instead of maintaining their course of generously overdimensioning their networks, operators have to drive their networks closer to the limits while trying to delay network expansion. This can be achieved, e.g., by replacing fixed-grid wavelength-division multiplexing (WDM) with a flexible-grid configuration or by reducing system margins during the network planning phase [2]. While these measures are purely technology-related, research proposes to take service level agreement (SLA)-related aspects into account as well. For example, the authors of [3] develop a model to estimate the availability of network services more precisely than with standard methods. In that way, the availability “safety margin” a network operator typically includes during provisioning can be selected appropriately. The authors of [4] consider the tradeoff between SLA penalties and protection costs with the goal of postponing costly protection investments.

In this work, we focus on the accurate fulfillment of a connection service’s availability according to its SLA specifications. We propose a service admission and routing approach that allows different network services to share excess availability. As a result, the amount of network resources that has to be reserved for protection can be reduced. In that way, the availability specified in the SLAs is provided with less overfulfillment than with many other routing approaches. Overall, this leads to more efficient network operation and enables the operator to accommodate more services in the network. Our mechanism is applicable to dynamically arriving services in connection-oriented networking technologies, e.g. wavelength services, optical transport network (OTN) paths, or multi-protocol label switching (MPLS) tunnels.

In the following, we will first discuss some important aspects about availability in SLAs. Afterwards, we present related work in Section III. Section IV introduces our admission and routing approach consisting of a mathematical model to estimate the SLA compliance probability and an admission and

routing algorithm. Section V shows some illustrative numerical results of the proposed approach for different meshed core networks. Finally, we conclude the paper in Section VI.

## II. AVAILABILITY IN SLAS

In this section we will discuss two important aspects of availability in SLAs. First, the difference between the steady-state availability and the service availability (or interval availability, respectively), the latter being more relevant for SLAs. Second, the overfulfillment of availability due to topological or protection-related constraints.

Contracts for network services between the network operator and its customers typically include an SLA. Among other things, the SLA defines performance targets like latency, availability, and protection for the network service. In case the network operator cannot meet those performance targets the customer can claim penalty payments or refunds. Typically, availability and protection are two very important SLA aspects. While some SLAs guarantee the use of protection explicitly (e.g. [5]), others only stipulate an availability level  $\alpha_{\text{SLA}}$  (e.g. [6]). In the latter case, the network operator is free to select appropriate protection mechanisms to provide the specified availability. In the following, we will focus on this particular scenario.

A common challenge in network operation is the admission and routing of dynamically arriving connection service requests. A simple but established strategy for network operators to select a suitable route and appropriate protection mechanisms is based on the evaluation of the steady-state availability  $a$  of a network service (e.g. [7], [8]). The steady-state availability is the probability to find a system working at an arbitrary point in time. A common formula to compute the steady-state availability is

$$a = \frac{MTTF}{MTTF + MTTR} \quad (1)$$

where  $MTTR$  is the mean time to repair and  $MTTF$  is the mean time to failure, i.e., the mean time between repair and next failure. If a combination of route and protection can be found that fulfills the condition  $a \geq \alpha_{\text{SLA}}$ , the service is accepted and routed. However, it has been shown by several authors, e.g., in [3], [9], [10], that the use of the steady-state availability can lead to unexpectedly high SLA violation ratios.

The reason is that the steady-state availability considers an infinite amount of time, but an SLA considers a *billing cycle* of finite length  $T$ , which is usually one month [5], [6]. At the end of each billing cycle, the availability during that cycle is evaluated according to

$$A = \frac{T - X}{T} \quad (2)$$

where  $X$  is the accumulated downtime of the service during the billing cycle.  $A$  is known as *interval availability* and in the context of this work we refer to it as *service availability*. Due to the different time horizons,  $a$  and  $A$  are not necessarily equal in value. Furthermore, since network failures occur randomly and also the time it takes to put the failed network segment back into operation is fraught with uncertainty, the accumulated downtime  $X$  and, consequently, also the service availability  $A$  are random

variables (RVs). Therefore, it is practically impossible to provide the stipulated availability to each and every service deterministically ( $A \geq \alpha_{\text{SLA}}$  almost surely). Instead, the service availability fulfills the stipulated availability only with a certain probability  $P(A \geq \alpha_{\text{SLA}})$  which we will call *compliance probability* or just *compliance* in the following (other works refer to the complementary probability as *SLA violation risk*, e.g. [9], [11]). Costly operator efforts, like protection, high-quality components, or fast repair, can increase this probability. However, even the most costly efforts cannot ensure  $P(A \geq \alpha_{\text{SLA}}) = 1$ , and hence, there will still be SLA violations. Therefore, from an economical perspective, increasing the compliance probability at any price is not necessarily the optimal choice. Instead, a network operator has to balance costs for the provision of availability against expected SLA penalties and by that select a suitable target level  $f_i$  for the compliance probability.

Regardless of whether the operator selects a service's route and protection based on the steady-state availability or the compliance probability, a fundamental problem that arises is that a service does usually not meet the SLA level or the operator's compliance target with equality ( $a = \alpha_{\text{SLA}}$  or  $P(A \geq \alpha_{\text{SLA}}) = f_i$ ) but only with *overfulfillment* ( $a > \alpha_{\text{SLA}}$  or  $P(A \geq \alpha_{\text{SLA}}) > f_i$ ). The reason is that different candidates for route selection do not offer an arbitrarily fine granularity of availability and also protection cannot increase availability in a continuous way but only in discrete steps. As a consequence, the network operator permanently provides more availability than has been specified in the SLAs. The routing methodology we propose reduces this overfulfillment and thus allows a more efficient network operation.

## III. RELATED WORK

Our admission and routing approach is not the only one to employ availability sharing or to make use of availability margins.

The authors of [12] propose a protection mechanism in which a broken wavelength service that is close to an SLA violation can preempt a wavelength service that still has downtime budget left. In that way, excess availability of services with little or no downtime is used to reduce SLA violations. A similar approach is proposed in [13] where a network service with a high SLA violation risk can preempt backup resources of services with lower risk. The violation risk is represented by the so-called urgency level (UL), a metric that takes the remaining holding time, the remaining affordable downtime, and the penalty costs of the service into account. The UL is also used in [14] where protection schemes (dedicated link protection, 1+N protection, and shared path protection) are changed dynamically to provide a suitable availability level on the one hand and to delay network upgrades on the other hand. A very similar approach using steady-state availability instead of UL is presented in [15]. In [16], the goal is the maximization of the overall profitability, i.e., the tradeoff between expected SLA penalties and service returns plays a key role during the service admission process. Similar to our approach, the authors employ the SLA violation probability instead of the steady-state

availability to find a suitable route. Also, the selected route is not required to be the one with the lowest SLA violation probability if this results in a higher total profitability. In that way, availability margins are exploited to improve profitability. The authors of [17] present a rather radical algorithm in which the network resources of an existing service are released for new services when the remaining holding time of the existing service is less than its remaining allowed downtime, i.e., when an SLA violation is no longer possible. The approach in [18] implements the idea of sharing excess availability by forming clusters of network services with heterogeneous SLAs. Services inside a cluster that have experienced only little downtime act as protection for services that are close to violating their SLA. In that way, no explicit backup resources are required. Finally, in [19]–[21], the authors propose an admission and routing approach for shared-path protection. Whenever a new service request arrives, the target availability of each existing service is recomputed, taking its remaining holding time and outage history into account. In that way, the target availability of a service with only little or no downtime is relaxed, and as a consequence, its potential sharing degree for shared-path protection is increased. For a service that already experienced much downtime, the target availability is increased and the potential sharing degree is decreased. Overall, it is shown that the adaptation of the target availabilities leads to less availability overfulfillment and a lower blocking ratio. [21] also discusses a possible way to employ the SLA violation risk instead of the steady-state availability but no results are shown.

All discussed publications contain aspects of our work but to the best of our knowledge there is no publication that targets the accurate fulfillment of the service availability according to the SLA specifications and employs the sharing of SLA compliance probability or violation risk to achieve this.

#### IV. ADMISSION AND ROUTING APPROACH

##### A. Overview

Our approach is responsible for the admission and routing or the rejection of randomly arriving connection service requests. As mentioned in the previous section, the goal of the approach is the reduction of availability overfulfillment. For the reasons described above, we employ the compliance probability and not the steady-state availability as a decision criterion in our approach. We assume that the network operator has set an internal compliance target level  $f_t$ . Therefore, the probability of SLA compliance must be larger than or equal to  $f_t$ , however, equality is intended. More formally, the operator wants to achieve

$$P(A \geq \alpha_{\text{SLA}}) \geq f_t \quad (3)$$

where  $A$  is the RV defined in (2) describing the monthly service availability of the provisioned services. All routing decisions of the algorithm are, in principle, based on (3).

When a service request arrives, the compliance probability of potential routes can be computed based on link properties (see Section IV-B). As argued above, it is unlikely that a route can be found for a service request  $i$  such that its compliance probability  $f_i$  equals the operator target precisely. Instead, there

TABLE I

EXEMPLARY SEQUENCE OF SERVICE REQUESTS AND SURPLUS SHARING. THE TARGETED COMPLIANCE LEVEL IS  $f_t = 0.99$  AND THE INITIAL SURPLUS IS  $\Delta f_0 = 0$ .

Request number $i$	Required compliance $f_{\text{req},i} = f_t - \Delta f_{i-1}$	Route compliance $f_i$	Resulting surplus $\Delta f_i = f_i - f_{\text{req},i}$
1	0.990	0.995	0.005
2	0.985	0.986	0.001
3	0.989	0.989	0.000

will be some overfulfillment  $\Delta f_i$ . In our routing algorithm, we accumulate the overfulfillment of individual services stepwise as *surplus* and share it with other services to reduce the overall overfulfillment. The sharing is realized by dynamically relaxing the compliance target for future service requests based on the stepwise accumulated surplus, i.e., instead of requiring a compliance probability of at least  $f_t$  for each and every service request, the  $i$ -th service request can be routed with a lower required compliance probability of

$$f_{\text{req},i} = f_t - \Delta f_{i-1} \quad (4)$$

where

$$\Delta f_{i-1} = f_{i-1} - f_{\text{req},i-1} \quad (5)$$

is the surplus accumulated until the previous service  $i - 1$ , and  $f_{i-1}$  is its compliance probability (or more precisely, that of the underlying route).

Table I shows an example with three consecutive service requests. The operator's target level is set to  $f_t = 0.99$  and the initial surplus  $\Delta f_0$  is zero. For the first request, potential routes must have a compliance probability of at least  $f_{\text{req},1} = 0.99$ . Next, our routing algorithm tries to find a suitable route. Assume that the route the algorithm selects has a compliance probability of  $f_1 = 0.995$ . The resulting difference is the surplus, namely  $\Delta f_1 = 0.995 - 0.99 = 0.005$ . This surplus is used to relax the compliance requirement of the second request. A route with a compliance probability of  $f_{\text{req},2} = 0.985$  is sufficient for the second request. The operator can benefit from this by selecting a less reliable route or by provisioning less protection.

As already mentioned, the algorithm needs to estimate the probability of SLA compliance for routes through the network. The mathematical model for this estimation is introduced in the next section. Afterwards, the details of the algorithm are presented.

##### B. Estimation of Compliance Probability

In this section, the estimation of the compliance probability for repairable services is introduced. A service can be in a working or a failure state. If a service fails, it will be repaired and is then working again. As in many other works, we assume that the time it takes to repair a service and the time until the next failure occurs both follow exponential distributions with means  $MTTR$  and  $MTTF$ , respectively. Based on (2), the availability  $A$

is a function of the cumulated downtime  $X$  during a billing cycle  $T$ . Therefore, the compliance probability can be expressed as

$$P(A \geq \alpha_{\text{SLA}}) = P(X \leq T \cdot (1 - \alpha_{\text{SLA}})) \quad (6)$$

$$= P(X \leq x_{\text{SLA}}) \quad (7)$$

where  $x_{\text{SLA}}$  is the maximum allowed cumulated service downtime per billing cycle  $T$ . The cumulative distribution function (CDF) of  $X$  has been derived in [22] and is given by

$$F(x) = a \cdot \Omega_{\lambda, \mu}(x) + (1 - a)(1 - \Omega_{\mu, \lambda}(T - x)) \quad (8)$$

where  $a = MTTF / (MTTF + MTTR)$  is the steady-state availability of the service,  $\lambda = 1/MTTF$ , and  $\mu = 1/MTTR$ . The function  $\Omega_{\gamma, \delta}(z)$  is the CDF of the RV  $Z$ . Like  $X$ ,  $Z$  describes the cumulated downtime during a time interval. However, in contrast to  $X$ , the service must be working at the beginning of the time interval. According to [23], we have

$$\Omega_{\gamma, \delta}(z) = e^{-\gamma(T-z)} \left( 1 + \sqrt{\gamma\delta(T-z)} \int_0^z e^{-\delta y} y^{-\frac{1}{2}} I_1 \left( 2\sqrt{\gamma\delta(T-z)y} \right) dy \right) \quad (9)$$

where  $I_1(x)$  is the modified Bessel function of the first kind of order 1, and  $\gamma$  and  $\delta$  correspond to the failure and repair rates of the process underlying  $Z$ . Finally, we obtain for the compliance probability

$$P(A \geq \alpha_{\text{SLA}}) = F(T \cdot (1 - \alpha_{\text{SLA}})). \quad (10)$$

In this work, we consider unprotected services and services with dedicated path protection, both of which consist of several network components. Therefore, in order to use (10), we have to aggregate the failure rates  $\lambda_j$  and repair rates  $\mu_j$  of the underlying components.

An unprotected connection service consists of a series of  $N$  network components. The service is working only if all components are working. We assume that component failures and repairs are independent of each other. Therefore, the aggregated failure rate  $\lambda_U$  is the sum of the individual failure rates [24, Ch. 6.3]

$$\lambda_U = \sum_{j=1}^N \lambda_j. \quad (11)$$

The steady-state availability of a series is given by the product of the individual availabilities, i.e.,

$$a_U = \prod_{j=1}^N a_j \quad (12)$$

and with (1) the repair rate of the service is

$$\mu_U = \lambda_U \frac{a_U}{1 - a_U}. \quad (13)$$

Dedicated path protection employs two parallel, disjoint routes: a working route and a backup route. Aggregating the two routes individually using the serial substitution above

yields  $\lambda_1$  and  $\mu_1$  for the working route and  $\lambda_2$  and  $\mu_2$  for the backup route. Then, according to [24, Ch. 6.4], the failure rate of such a protected service is

$$\lambda_P = \frac{\lambda_1 \lambda_2 (\lambda_1 + \lambda_2 + \mu_1 + \mu_2)}{(\lambda_1 + \mu_2)(\lambda_2 + \mu_1) + \lambda_1(\lambda_1 + \mu_2) + \lambda_2(\lambda_2 + \mu_1)}. \quad (14)$$

The steady-state availability of the two routes is given by

$$a_P = 1 - (1 - a_1)(1 - a_2) \quad (15)$$

and similar to (13) the repair rate is

$$\mu_P = \lambda_P \frac{a_P}{1 - a_P}. \quad (16)$$

### C. Admission and Routing Algorithm

The routing algorithm handles a service request by accepting and routing it if enough network capacity is available and a route with sufficient compliance can be found. If this is not the case, the service request is blocked. The algorithm has a global view on the network, i.e., it knows all link capacities and the current traffic on the links as well as the link failure and repair rates. The network is represented by a graph  $G$  comprising the nodes and links. A service request  $R$  consists of a source and a destination node, a data rate, and an SLA availability. We assume that a service cannot be split but must be served by a single route.

Algorithm 1 shows how an arriving service request is handled. First, a subgraph of  $G$  is created which only contains links that are currently working and have enough free capacity to carry the new service. We assume that network nodes are failure-free. In line 3, the required compliance probability for the service route is determined. As explained above, our algorithm uses surplus sharing, i.e., the required compliance is the operator's target compliance  $f_t$  relaxed by the current surplus  $\Delta f$ . The case without surplus sharing serves as a reference in the evaluation below. Next, the algorithm searches for a suitable route. If the search succeeds, the service is routed accordingly and the accumulated surplus is set to the current route's surplus. Otherwise, the service is blocked.

Algorithm 2 shows how a suitable route is found. Essentially, the algorithm first searches for an unprotected primary route with enough compliance probability using the shortest path in the subgraph  $G_s$ . If the unprotected route is not reliable enough, a link-disjoint backup route is added and the compliance is evaluated again (lines 4–16). The whole procedure is repeated up to  $k_{\text{max}}$  times with alternative primary paths (line 3). This is important for two reasons. First, depending on the shortest path metric, the shortest path is not necessarily the path with the highest compliance probability. As a result, a  $k$ -th shortest path with  $k > 1$  could provide sufficient compliance while the first shortest path does not. Second, if  $G_s$  is rather sparse, the bad choice of a primary path can lead to the situation that no additional backup path is available. The compliance probability in lines 8 and 14 is calculated using the procedure presented in Section IV-B.

**Algorithm 1** Admission and routing.**Global state**

- Network graph  $G$
- Operator compliance target  $f_t \in [0, 1]$
- Cumulated compliance surplus  $\Delta f \geq 0$

**Input**

Service request  $R = (s, d, h, \alpha_{\text{SLA}})$ , with source node  $s$ , destination node  $d$ , data rate  $h$ , and SLA availability  $\alpha_{\text{SLA}}$

```

1 procedure HANDLESERVICEREQUEST( $R$ )
2    $G_s \leftarrow$  subgraph of  $G$  including only links
   that are working and have free capacity  $\geq h$ 
3    $f_{\text{req}} \leftarrow f_t - \Delta f$  if surplus sharing enabled else  $f_t$ 
4    $route, f \leftarrow$  FINDROUTE( $G_s, R, f_{\text{req}}$ )
5   if  $route \neq \text{None}$  then
6     Route service request on  $route$ 
7      $\Delta f \leftarrow f - f_{\text{req}}$ 
8   else
9     Block service request

```

**Algorithm 2** Route selection.**Global state**

Maximum primary path trials  $k_{\text{max}} \in \mathbb{N}^+$

**Input**

- Network subgraph  $G_s$
- Service request  $R = (s, d, h, \alpha_{\text{SLA}})$
- Required compliance  $f_{\text{req}} \in \mathbb{R}$

**Output**

Feasible route and its compliance probability or None

```

1 function FINDROUTE( $G_s, R, f_{\text{req}}$ )
2    $k \leftarrow 1$ 
3   while  $k \leq k_{\text{max}}$  do
4      $p \leftarrow$  KTHSHORTESTPATH( $G_s, s, d, k$ )
5     if  $p = \text{None}$  then
6       return None
7     else
8        $f \leftarrow$  COMPLIANCEPROB( $p, \alpha_{\text{SLA}}$ )
9       if  $f \geq f_{\text{req}}$  then
10        return  $p, f$   $\triangleright$  Unprotected route
11       $G_b \leftarrow$  subgraph of  $G_s$  excluding links in  $p$ 
12       $b \leftarrow$  SHORTESTPATH( $G_b, s, d$ )
13      if  $b \neq \text{None}$  then
14         $f \leftarrow$  COMPLIANCEPROB( $(p, b), \alpha_{\text{SLA}}$ )
15        if  $f \geq f_{\text{req}}$  then
16          return  $(p, b), f$   $\triangleright$  Protected route
17       $k \leftarrow k + 1$ 
18 return None

```

## V. ILLUSTRATIVE NUMERICAL EXAMPLE

## A. Simulation Setup

We evaluate the performance of our routing approach in four different core networks. The networks are taken from [25]–[27] and are depicted in Figure 1. Table II provides additional information. Based on realistic values provided in [28], we

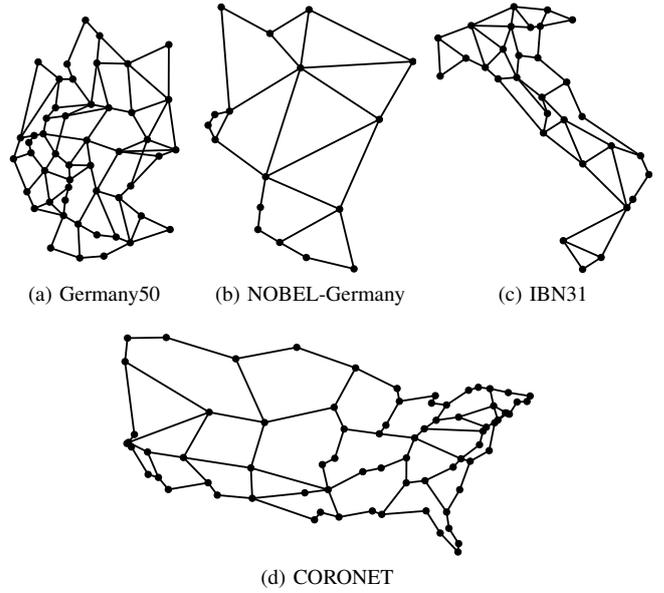


Fig. 1. Network topologies.

TABLE II  
NETWORK PROPERTIES.

Network	Nodes	Links	Link length in km		
			Min.	Avg.	Max.
Germany50	50	88	26	101	252
NOBEL-Germany	17	26	29	143	294
IBN31	31	52	34	128	283
CORONET	75	99	20	330	1017

assume an MTTR of 9 hours for all links, and an MTTF of

$$MTTF_e = \frac{628 \text{ km} \cdot 360 \text{ days} \cdot 24 \text{ h/day}}{\ell_e} \quad (17)$$

for link  $e$ , where  $\ell_e$  is the link length. We assume that nodes do not fail. For the sake of simplicity, a year in the simulation has 360 days with 30 days per month. Each link provides a capacity of 16 Tbit/s. Service requests arrive randomly with an exponentially distributed inter-arrival time (IAT). The mean IAT is varied over different simulation runs to evaluate different network loads. The source-destination node pair of a service request is chosen uniformly from all node pairs in the network. The requested data rate and holding time (contract period) are selected uniformly from 40 or 100 Gbit/s and 3, 6, 12, or 24 months, respectively. The parameter  $k_{\text{max}}$  is set to 5. The weight of a link  $e$ , used by the shortest path algorithm, is  $-\log(a_e)$ , where  $a_e$  is the steady-state availability of the link. This link weight allows a standard shortest path algorithm to find the path with the highest availability—the so-called *most reliable path* [29]. In general, the path with the highest availability is not necessarily the same as the path with the highest compliance probability<sup>1</sup>. However,

<sup>1</sup>For example, with  $T = 1$  month,  $\alpha_{\text{SLA}} = 0.99999$ ,  $MTTR_1 = 7$  h, and  $a_1 = 0.9999$ , the compliance probability is  $f_1 = 0.9897$ , while with  $MTTR_2 = 15$  h, and  $a_2 = 0.9998 < a_1$  it is  $f_2 = 0.9903 > f_1$ , due to the change in the MTTR.

in practice, they often coincide. We do not assume a specific networking technology in this evaluation, therefore, typical challenges, especially from the optical domain, e.g. spectrum fragmentation, contiguity, and continuity, are not considered.

We study the behavior of our algorithm for compliance targets  $f_t$  of 0.99, 0.995, and 0.999. As an example, for  $f_t = 0.99$ , the operator only allows 1% of all billing cycles to violate the SLA availability. For the networks Germany50, NOBEL-Germany, and IBN31, the SLA availability is set to  $\alpha_{\text{SLA}} = 0.99999$ . For the network CORONET,  $\alpha_{\text{SLA}} = 0.9828$ . With those availability levels, the feasibility of a service is guaranteed for all nodes pairs, even with the most strict compliance target of 0.999.

For each set of parameters, 500 consecutive periods (batches) of 5 years each have been simulated in an event-based batch simulation using the IKR SimLib library [30]. The startup phase in each simulation run has been set to 100 years to eliminate transient effects.

## B. Results

We first discuss detailed results for the Germany50 network with a compliance target of  $f_t = 0.99$ . Figure 2 shows plots for the protection overbuild, the compliance ratio, and the blocking ratio. The protection overbuild is the number of links in the protection path divided by the number of links in the primary path. In the case of an unprotected route, the overbuild is zero. For each plot, the *load factor* is varied to evaluate the algorithm under different network loads. The load factor is varied by changing the mean IAT of service requests. The mean IAT has been calibrated per parameter combination in a preparatory simulation such that the blocking ratio for a load factor of 1 is at around  $10^{-2}$  when surplus sharing is disabled. The obtained mean IATs range from 7 h to 16 h. The simulation with surplus sharing enabled uses the same calibration. The error bars in Figure 2c depict 95 %-confidence intervals for the mean. In Figures 2a and 2b, the confidence intervals are omitted because they are too small to be visible.

Figure 2a shows the average protection overbuild. Using surplus sharing, the protection overbuild is reduced by around 38% because some services can be routed without protection due to the accumulated surplus. With increasing network load, the overbuild increases as well because the shortest paths to connect a node pair are more and more occupied and alternate paths have to be taken.

Figure 2b shows the compliance ratio. The compliance ratio in a simulation batch describes the share of all billing cycles in which the service availability fulfilled the SLA. It is calculated based on all provisioned services and their billing cycles as

$$\frac{1}{\sum_{i=1}^N M_i} \sum_{i=1}^N \sum_{m=1}^{M_i} \mathbf{1}_{a_{i,m} \geq \alpha_{\text{SLA}}} \quad (18)$$

where  $N$  is the number of provisioned services,  $M_i$  is the number of billed months of the  $i$ -th service (3, 6, 12, or 24) and  $a_{i,m}$  is its actual service availability in its  $m$ -th billed month, i.e., the realization of the RV  $A$  in (2).  $\mathbf{1}_{a_{i,m} \geq \alpha_{\text{SLA}}}$  is the

indicator function which equals 1 if  $a_{i,m} \geq \alpha_{\text{SLA}}$  (availability high enough) and 0 otherwise. The connected marks in the figure show the average compliance ratio over all 500 batches. The medium-sized marks correspond to the first decile and the small marks represent the minimum 5-year compliance ratio of the 500 batches. Without surplus sharing, the average compliance ratio is close to 1 which means that the operator target of  $f_t = 0.99$  is overfulfilled. With surplus sharing enabled, the average compliance ratio is close to 0.99. Consequently, the overfulfillment has been reduced significantly. The average ratios are independent of the network load. Since the SLA compliance is the result of a random failure and repair process, there are periods in which the compliance ratio is below the targeted level. As can be seen, of the 500 5-year periods, the minimum compliance ratios with surplus sharing are around 0.986 while the first deciles are around 0.989. Of course, these variations pose a risk for the network operator which should be minimized. The amount of variation depends on the failure characteristics of the network components and on the service arrival and holding time behavior. Furthermore, the considered time period plays an important role, i.e., time horizons longer than the currently considered 5 years will result in less variation. However, the detailed relations will be the subject of future work.

Figure 2c shows the average blocking ratio. It can be seen that surplus sharing leads to consistently lower blocking ratios. Considering the network load at a blocking ratio of  $10^{-2}$ , surplus sharing allows a significant increase of about 11% (solid ellipse in Figure 2c). For low-load situations (load factor  $\leq 0.8$ ), the blocking ratio is reduced by around 15% (dotted ellipse in Figure 2c).

To summarize, the proposed admission and routing approach is able to almost eliminate the availability overfulfillment, it reduces the resource overbuild required for protection considerably, and it allows more services to be accommodated in the Germany50 network. We will now include results for the other networks and stricter compliance targets of 0.995 and 0.999. For the protection overbuild and the compliance ratio, the relative change when enabling surplus sharing is almost independent from the load factor. Therefore, we show values averaged over all load factors in the following. Furthermore, we show the attainable load increase at a blocking ratio of  $10^{-2}$  but do not consider the change in the blocking ratio for low network loads as we did above. Notice that in each of the following figures the bar for Germany50 and  $f_t = 0.99$  relates to the values presented in Figure 2.

Figure 3 shows the reduction in protection overbuild averaged over all load factors. The behavior is very similar for all networks. The highest reduction of more than 36% is possible for a compliance target of  $f_t = 0.99$ . For stricter targets the reductions decrease. Nevertheless, a significant reduction of around 10% is possible in all networks even for a compliance target of  $f_t = 0.999$ .

Figure 4 compares the compliance ratio averaged over all load factors with and without surplus sharing. Figure 4a depicts the case without surplus sharing. It is visible that the resulting compliance ratio is above 0.999 for all networks

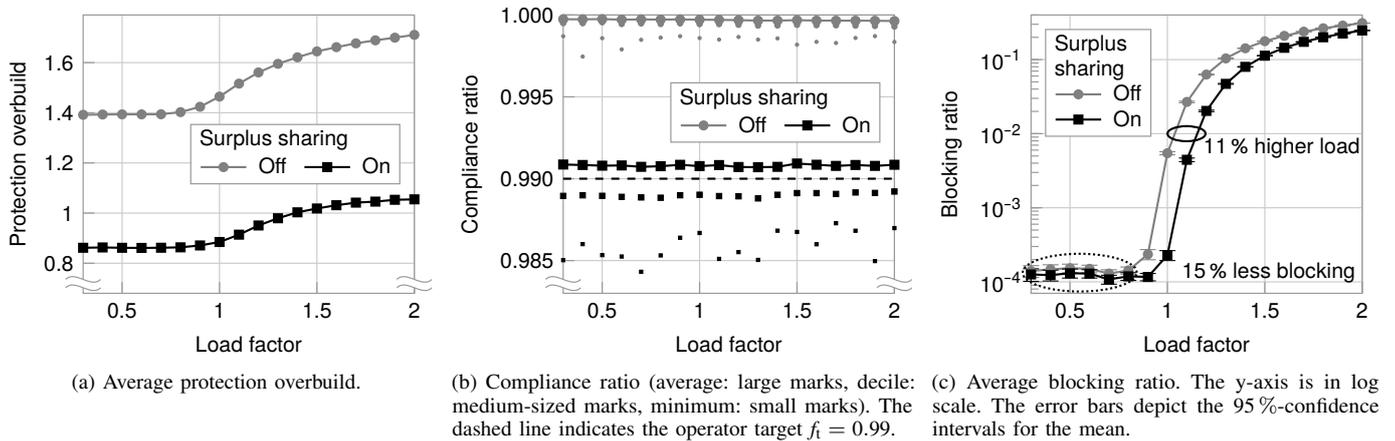


Fig. 2. Results for the Germany50 network with a compliance target of  $f_t = 0.99$  for various network loads.

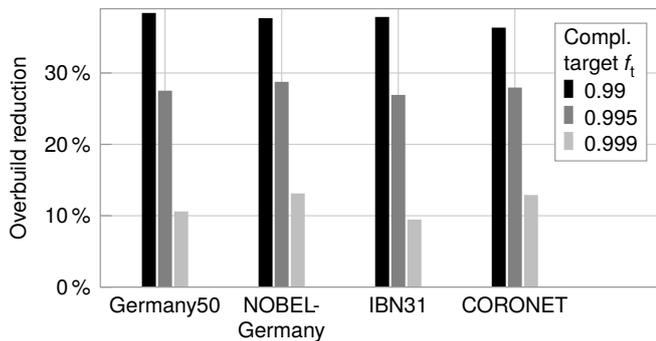


Fig. 3. Reduction in protection overbuild.

almost independently of the actual compliance target in use. Consequently, among the scenarios in this study, the availability overfulfillment is the highest for a compliance target of  $f_t = 0.99$ . As mentioned earlier, the high overfulfillment mainly originates from the coarse availability granularity of the protection—a backup path raises the compliance probability considerably even though a small increase might have been sufficient. Figure 4b shows the compliance ratio with surplus sharing enabled. It can be seen that the overfulfillment is reduced significantly in all networks. For the CORONET network, the compliance ratio matches the compliance target tightly. For the other networks, a certain margin persists.

Finally, Figure 5 shows the increase in attainable network load for a blocking ratio of  $10^{-2}$ . The highest increase is achieved for a target compliance of 0.99. Stricter compliance targets result in less load increase. To a certain degree, the load increase is related to the overfulfillment reduction. Stronger overfulfillment reductions (see Figure 4) yield stronger increases in network load. Except for the network IBN31, load increases of more than 11% for a target compliance of 0.99 are achieved by our admission and routing approach. For a target compliance of 0.995, load increases between 3.6% and 7.9% are realized.

The results show that the approach we propose is able to substantially increase the amount of services a network operator can accommodate in its network.

## VI. CONCLUSION

In order to improve network efficiency, operators have to reduce margins that are present in the network. In this work, we have identified availability overfulfillment as a type of margin that appears unintentionally due to topological or protection-related constraints. Since availability overfulfillment ties up valuable network resources, an overfulfillment reduction can improve the network efficiency.

We have proposed a service admission and routing approach that is able to reduce the availability overfulfillment by sharing surplus availability among services. More precisely, our algorithm considers the probability of SLA compliance of connection services with and without dedicated path protection. Excess compliance probability of one service is accumulated as surplus, which is used to relax availability requirements of other services.

The simulation results confirm that the availability overfulfillment is reduced or even eliminated, and that fewer resources have to be dedicated to protection. As a result, the blocking ratio of new connection services is reduced significantly and more services can be accommodated in the network. The results also show that, even though the average compliance ratio matches the operator target accurately, the compliance ratio can fall below the target level for certain periods of time due to statistical variations. Further work is required to incorporate this behavior into the overall model. Nevertheless, the presented approach is a powerful mechanism to exploit existing margins in the network and to improve network efficiency.

## REFERENCES

- [1] P. J. Winzer and D. T. Neilson, "From Scaling Disparities to Integrated Parallelism: A Decathlon for a Decade," *Journal of Lightwave Technology*, vol. 35, no. 5, pp. 1099–1115, 2017.
- [2] Y. Pointurier, "Design of Low-Margin Optical Networks," *Journal of Optical Communications and Networking*, vol. 9, no. 1, pp. A9–A17, 2017.
- [3] L. Zhou and W. D. Grover, "A Theory for Setting the "Safety Margin" on Availability Guarantees in an SLA," in *5th International Workshop on Design of Reliable Communication Networks (DRCN)*. IEEE, 2005.
- [4] C. Meusburger and D. A. Schupke, "Method to Estimate the Break-Even Point Between SLA Penalty Expenses and Protection Costs," in *2008 Optical Fiber Communication Conference (OFC)*. Optical Society of America, 2008.

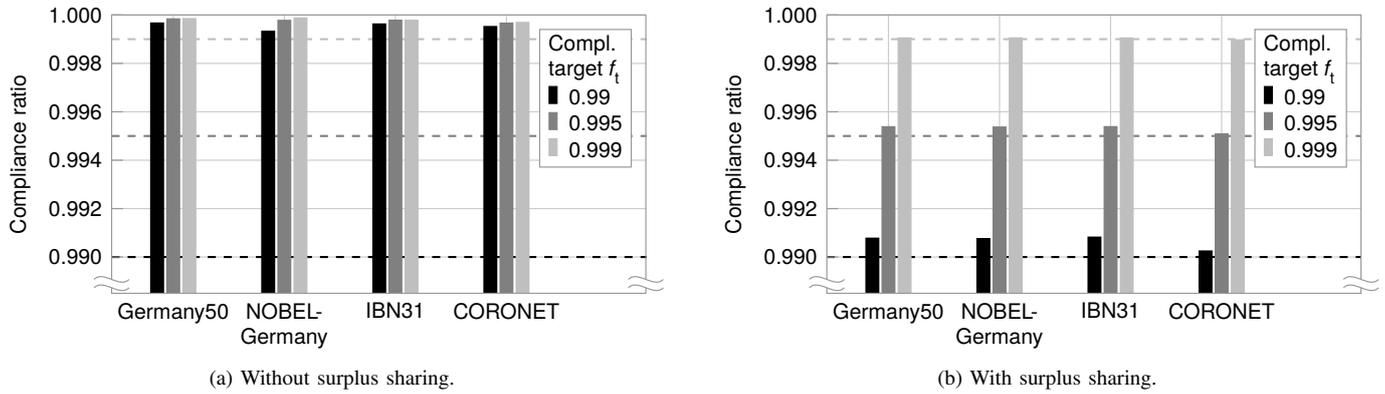


Fig. 4. Compliance ratio. The dashed lines correspond to the targeted compliance probabilities.

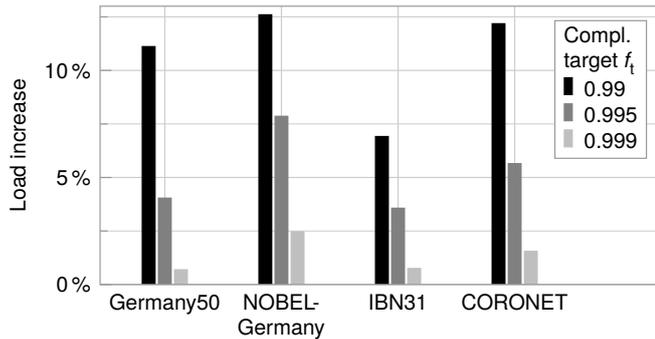


Fig. 5. Increase in attainable network load for a blocking ratio of  $10^{-2}$ .

- [5] Verizon Communications Inc., “WAVELENGTH SERVICES SOLUTION +; Part IV: Service Level Agreement,” [https://www.verizon.com/business/service\\_guide/reg/cp\\_wss\\_plus\\_sla.pdf](https://www.verizon.com/business/service_guide/reg/cp_wss_plus_sla.pdf), 2020, Online, accessed 31 March 2022.
- [6] Windstream Enterprise, “COMPLETE DATA MPLS: SERVICE LEVEL AGREEMENT,” [https://www.windstreamenterprise.com/wp-content/uploads/legal/MPLS\\_SLA.pdf](https://www.windstreamenterprise.com/wp-content/uploads/legal/MPLS_SLA.pdf), 2019, Online, accessed 31 March 2022.
- [7] M. Tornatore, G. Maier, and A. Pattavina, “Availability Design of Optical Transport Networks,” *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 8, pp. 1520–1532, 2005.
- [8] L. Song, J. Zhang, and B. Mukherjee, “Dynamic Provisioning with Availability Guarantee for Differentiated Services in Survivable Mesh Networks,” *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 3, pp. 35–43, 2007.
- [9] M. Xia, J. ho Choi, and T. Wang, “Risk Assessment in SLA-Based WDM Backbone Networks,” in *2009 Optical Fiber Communication Conference (OFC)*. Optical Society of America, 2009.
- [10] A. J. González and B. E. Helvik, “Guaranteeing Service Availability in SLAs: a Study of the Risk Associated with Contract Period and Failure Process,” *IEEE Latin America Transactions*, vol. 8, no. 4, pp. 410–416, 2010.
- [11] L. Mastroeni and M. Naldi, “Violation of Service Availability Targets in Service Level Agreements,” in *2011 Federated Conference on Computer Science and Information Systems (FedCSIS)*. IEEE, 2011.
- [12] O. Gerstel and G. Sasaki, “Meeting SLAs by Design: a Protection Scheme with Memory,” in *2007 Optical Fiber Communication Conference (OFC)*. Optical Society of America, 2007.
- [13] M. Xia, M. Batayneh, L. Song, C. U. Martel, and B. Mukherjee, “SLA-Aware Provisioning for Revenue Maximization in Telecom Mesh Networks,” in *2008 IEEE Global Telecommunications Conference (GLOBECOM)*. IEEE, 2008.
- [14] F. Dikbiyik, M. Tornatore, and B. Mukherjee, “Exploiting Excess Capacity, Part II: Differentiated Services Under Traffic Growth,” *IEEE/ACM Transactions on Networking*, vol. 23, no. 5, pp. 1599–1609, 2015.
- [15] X. Chen, M. Tornatore, S. Zhu, F. Ji, W. Zhou, C. Chen, D. Hu, L. Jiang, and Z. Zhu, “Flexible Availability-Aware Differentiated Protection in Software-Defined Elastic Optical Networks,” *Journal of Lightwave Technology*, vol. 33, no. 18, pp. 3872–3882, 2015.
- [16] A. Das, “Maximizing Profit Using SLA-Aware Provisioning,” in *2012 IEEE Network Operations and Management Symposium (NOMS)*. IEEE, 2012.
- [17] A. Nafarieh, S. C. Sivakumar, W. Phillips, and W. Robertson, “Memory-aware SLA-based Mechanism for Shared-Mesh WDM Networks,” in *3rd International Congress on Ultra Modern Telecommunications and Control Systems (ICUMT)*. IEEE, 2011.
- [18] M. Xia, C. U. Martel, M. Tornatore, and B. Mukherjee, “Service Cluster: A New Framework for SLA-Oriented Provisioning in WDM Mesh Networks,” in *2009 IEEE International Conference on Communications (ICC)*. IEEE, 2009.
- [19] M. Tornatore, D. Lucerna, L. Song, B. Mukherjee, and A. Pattavina, “Dynamic SLA Redefinition for Shared-Path-Protected Connections with Known Duration,” in *2008 Optical Fiber Communication Conference (OFC)*. Optical Society of America, 2008.
- [20] D. Lucerna, M. Tornatore, B. Mukherjee, and A. Pattavina, “Availability Target Redefinition for Dynamic Connections in WDM Networks with Shared Path Protection,” in *7th International Workshop on Design of Reliable Communication Networks (DRCN)*. IEEE, 2009.
- [21] —, “Trading availability among shared-protected dynamic connections in WDM networks,” *Computer Networks*, vol. 56, no. 13, pp. 3150–3162, 2012.
- [22] T. Enderle, “On the Impact of Billing Cycles on Compensations in Network SLAs,” in *25th International Conference on Optical Network Design and Modeling (ONDM)*. IEEE, 2021.
- [23] L. Takács, “ON CERTAIN SOJOURN TIME PROBLEMS IN THE THEORY OF STOCHASTIC PROCESSES,” *Acta Mathematica Academiae Scientiarum Hungarica*, vol. 8, pp. 169–191, 1957.
- [24] A. Birolini, *Reliability Engineering: Theory and Practice*, 8th ed. Berlin, Germany: Springer-Verlag, 2017.
- [25] S. Orłowski, R. Wessály, M. Pióro, and A. Tomaszewski, “SNDlib 1.0—Survivable Network Design Library,” *Networks: An International Journal*, vol. 55, no. 3, pp. 276–286, 2010.
- [26] A. Eira, J. Pedro, and J. Pires, “Optimized Design of Shared Restoration in Flexible-Grid Transparent Optical Networks,” in *2012 Optical Fiber Communication Conference (OFC)*. Optical Society of America, 2012.
- [27] A. Von Lehmen, R. Doverspike, G. Clapp, D. M. Freimuth, J. Gannett, A. Kolarov, H. Kobrinski, C. Makaya, E. Mavrogiorgis, J. Pastor, M. Rauch, K. K. Ramakrishnan, R. Skoog, B. Wilson, and S. L. Woodward, “CORONET: Testbeds, Demonstration, and Lessons Learned [Invited],” *Journal of Optical Communications and Networking*, vol. 7, no. 3, pp. A447–A458, 2015.
- [28] S. Verbrugge, D. Colle, P. Demeester, R. Huelsermann, and M. Jaeger, “General Availability Model for Multilayer Transport Networks,” in *5th International Workshop on Design of Reliable Communication Networks (DRCN)*. IEEE, 2005.
- [29] J. Zhang, K. Zhu, H. Zang, and B. Mukherjee, “A New Provisioning Framework to Provide Availability-Guaranteed Service in WDM Mesh Networks,” in *2003 IEEE International Conference on Communications (ICC)*. IEEE, 2003.
- [30] J. Sommer and J. Scharf, “IKR Simulation Library,” in *Modeling and Tools for Network Simulation*, K. Wehrle, M. Güneş, and J. Gross, Eds. Springer, Berlin, Heidelberg, 2010, pp. 61–68.

# Towards a Hybrid Architecture by Introducing Coherent Pluggable Transceivers in IP-Optical Core Networks with Optical Cross-Connects

Filippos Christou

*Institute of Communication Networks  
and Computer Engineering (IKR)  
University of Stuttgart  
Stuttgart, Germany  
filippos.christou@ikr.uni-stuttgart.de*

Tobias Enderle

*Institute of Communication Networks  
and Computer Engineering (IKR)  
University of Stuttgart  
Stuttgart, Germany  
tobias.enderle@ikr.uni-stuttgart.de*

Arthur Witt

*Institute of Communication Networks  
and Computer Engineering (IKR)  
University of Stuttgart  
Stuttgart, Germany  
arthur.witt@ikr.uni-stuttgart.de*

**Abstract**—During the previous years, there has been a big technological advancement in Digital Signal Processing (DSP) and coherent pluggable optics, which enabled a considerable increase in the data rate and a step towards economical packet-optical integration, respectively. This caused the network operators to start wondering whether an architectural change should take place in their infrastructure. Coherent Pluggable Transceivers (CPTs), already being a success story for Data Center Interconnect (DCI), have grown to be used not only in metro networks but, as of recently, also in some parts of the core. The latter constitutes a major use case for the Multi-Source Agreement (MSA) OpenZR+. As many network operators can have good reasons to stick with the traditional standalone transponder equipment, this paper explores a hybrid solution combining CPTs and transponders to fully exploit the advantages of the two technologies. More specifically, we balance the trade-off between using cheaper pluggable modules directly on the IP router and powerful standalone flexible transponders with a gray interface. Although the integration of colored optics in the router was already proposed about a decade ago with Cisco's IPoD-WDM, we find that this time it is more appealing in terms of power consumption, module size, and equipment cost. Furthermore, we look closer at a collapsed packet-optical integrated architecture of Hop-by-Hop lightpath connections as opposed to a multilayer one, which also uses Optical Cross-Connects (OXC)s. We show that a hybrid solution using both architectures yields the best results for improving network metrics and minimizing hardware costs. We contribute an Integer Linear Program (ILP) to solve the dimensioning problem for realistic countrywide topologies.

**Index Terms**—Coherent pluggable transceivers, Hybrid architecture, Integer linear program, Network dimensioning

## I. INTRODUCTION

For decades, telecommunication networks have faced increasing traffic demands, and they will continue to do so in the foreseeable future. Today, transporting all this traffic without coherent optical transmission is hard to imagine. The market for optical transmission equipment is highly competitive. Network vendors are constantly improving their transponders to achieve higher signal reaches and larger capacities using sophisticated

modulation schemes like Probabilistic Constellation Shaping (PCS), advanced Forward Error Correction (FEC), multi-carrier super-channels, and higher baud rates. In the following, we refer to this class of transponders as Coherent Elastic Transponders (CETs). While the evolution of CETs is important for transport networks of large telecommunication providers, coherent optics also found their way into the Data Center (DC) ecosystem during the last years. Inside the DC, traffic volumes have been increasing even stronger than in traditional transport networks, and the interconnection of such DCs now requires coherent optics as well. However, in contrast to the traditional telecommunication providers' approach of using top-notch, proprietary CETs, DC operators tend to focus on simplicity, cost efficiency, and interoperability. As a result, the Optical Internetworking Forum (OIF) released the 400ZR Implementation Agreement (IA), a de facto standard for a Coherent Pluggable Transceiver (CPT) for 400 Gbps Ethernet-based transmission. Allowing only 400 Gbps and no other data rates at a reach of about 120 km, 400ZR transceivers are less powerful and less flexible than CETs. However, less flexibility means easier handling, and since 400ZR is an IA, transceivers of one vendor are interoperable with those of other vendors, which provides advantages in scalability, disaggregation, and cost efficiency. Additionally, 400ZR transceivers benefit from the growing integration of both electrical and optical components, which allows networking equipment to be built much smaller and much more power-efficient than in the past. As a result, 400ZR transceivers come in the QSFP-DD form factor and can be plugged into the router or switch ports directly. This eliminates the need for separate optics shelves and gray connections required by traditional transponder-based setups.

Inspired by the success of 400ZR, the OpenZR+ Multi-Source Agreement (MSA) was developed [1]. A ZR+ is another CPT that supports different modulation formats, several data rates and boosts signal reach with a standardized FEC. With that, it can be considered a viable alternative to traditional transponder-based setups in metro or even core networks of telecommunication providers. The principle idea is not new.

This work has been performed in the framework of the CELTIC-NEXT EUREKA project AI-NET-ANTILLAS (Project ID C2019/3-3), and it is partly funded by the German BMBF (Project ID 16KIS1312). The authors alone are responsible for the content of the paper.

More than a decade ago, Cisco developed the IP over DWDM (IPoDWDM) technology with the goal of integrating optics into the router. However, the technology did not gain traction because IPoDWDM line cards were expensive and occupied many precious router slots. Furthermore, the strict separation among telecommunication providers between IP and optics departments made the integration difficult. Nowadays, the separation of departments appears to be less strict, and the cost and space characteristics of OpenZR+ are better than those of IPoDWDM. The first providers already use OpenZR+ transceivers in their core networks [2], but the majority seems still hesitant.

In this work, we compare different scenarios to decide whether OpenZR+ is a sensible choice for a provider's network. More importantly, we find the best way to combine CPTs and CETs in the considered problem instances by developing an Integer Linear Program (ILP). On top of that, we examine the influence of a Hop-by-Hop (HbH) architecture by reducing the number of Optical Cross-Connects (OXC) in the network and discover how much that can be beneficial. We realize that incorporating a combination of all these paradigms is the most cost-efficient yet powerful enough architecture case for the network dimensioning problem.

The following section introduces literature related to our study. In Section III, we put the problem statement in concrete terms. Section IV presents the assumed architecture model. Section V contains a detailed description of the ILP we used for network dimensioning. Section VI introduces the different scenarios we compared, and the corresponding results are discussed in Section VII. Finally, we conclude the paper in Section VIII.

## II. RELATED WORK

In [3] the British Telecommunications Group analyzed different network scenarios where CPTs are used. They also consider traditional Reconfigurable Optical Add-Drop Multiplexer (ROADM) networks with CETs to allow direct connections between source and destination nodes as a reference solution. An interesting scenario is a mixed network architecture, where 400 Gbps CPTs are used on links where the distance allows it, and traditional 400G CETs are used for the longer ones. The evaluation of an 8-year network planning scenario with a 30% traffic increase per year shows that the mixed architecture will be the most cost-efficient in the long run.

The authors of [4] introduced an ILP to compare CPTs with CETs under the consideration of line-side protection or optical restoration mechanisms. Line-side protection requires the provisioning of a protection path with the same channel settings as the working path. CETs can be used here very efficiently as only a few additional transponders are needed. CPTs will have to reduce the transmission rate to reach the longer distances for the protection paths, which will increase the number of modules significantly. In optical restoration, where spectral resources can be shared for working and protection paths, the number of CPT interfaces is greatly reduced.

IP-optical architectures for metro, regional, and long-haul networks are compared in [5]. The authors identify that CETs are suitable for long-haul networks as they provide the best spectral

efficiency and largest optical reach. For metro or regional networks, CPTs directly plugged in at router ports are preferable.

The most recent publication in this field [6] considers three network architectures: (a) CPTs in a HbH network, (b) CPTs with an OXC at every node, and (c) CETs with rates of up to 800 Gbps and an OXC at every node. Their ILP-based evaluation for the US Coronet and a pan-India topology significantly reduces required equipment like transmission modules and gray interfaces if the flexible and high-performance CETs are used.

## III. OBJECTIVE

Traditionally, the IP-optical core networks are equipped with high-end CETs that provide increased flexibility in adaptive modulation schemes and utilization of optical bandwidth. Nowadays, inter-DC connections are often realized with CPTs to offer an exclusive and direct connection between the data centers. These connections are integrated into the traditional core networks, which naturally leads to a network architecture with various transmission modules. Consequently, we want to discuss if a mixture of CETs and CPTs can be used to increase the efficiency of core networks while keeping low costs. Further, various network architectures will be studied. First, a traditional multilayer network that allows optical bypassing. Second, a flat HbH network. Third, a hybrid architecture where only some nodes are equipped with an OXC. This architecture aims mainly to reduce hardware costs and still provide the opportunity for bypassing.

Those issues can be seen as network planning and dimensioning problems that are highly interesting for network operators who want to keep their equipment up to date. Therefore, we want to introduce a method to optimize and evaluate network configurations with various transmission module technologies based on an ILP formulation. We also overview several metrics, such as router port utilization, spectrum allocation, signal regenerations, latency, energy consumption, and hardware costs. These metrics are evaluated for both cases of electrical grooming being allowed or not.

## IV. ARCHITECTURE MODEL

This section will describe some architectural choices and assumptions fundamental to carrying out the study as described in the previous chapter.

### A. Network and Network Nodes

Let us consider a network described by a unidirectional graph  $G(V, E)$  with  $V$  the set of vertices and  $E$  the set of edges. For each  $e \in E$ , we signify the distance of the fiber used on this link as  $length(e)$ . The set  $U = \{(u, v) \in V \times V : u \neq v\}$  contains all node pairs. Each vertex in the graph corresponds to a node in the network, either a router or a router and an OXC. Each router contains several line cards, where each line card  $l \in L$  has a tuple of properties  $(n_l, r_l, c_l)$  being the port density, the maximum port rate, and the cost of the line card, respectively. Having a set of line cards  $L$  and a specific line card  $l$ , we signify the line card with the next lower port rate in the set as  $l_{low} = lower(L, l)$ . Table I shows the available line cards  $L$ .

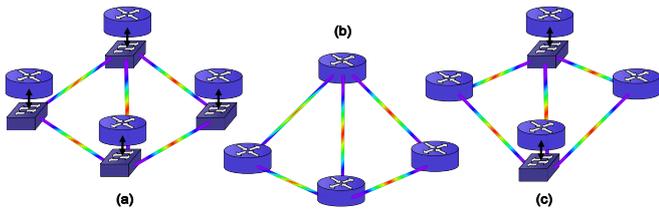


Fig. 1. The (a) Multilayer, (b) HbH, and (c) hybrid architectures.

The values are based on the predictions of [7], meaning that costs are normalized to the cost of a 10 Gbps transponder in the year 2012. Moreover, the line cards are grouped into line card chassis, with each one having a constant capacity of  $N_{\text{icc}} = 16$  line card slots and a price of 4.7 cost units. An OXC is realized using a colorless, directionless, contentionless (CDC) ROADM and serves to bypass the optical signals towards neighboring nodes without having to undergo Optical-Electrical-Optical (O-E-O) conversion. The percentage of OXCs in the network signifies the number of nodes having an OXC, and it can range from 0% being a full *HbH* architecture to 100% being a full *multilayer* architecture. Of great interest in this study will be all the values in between, and such architecture will be characterized as *hybrid*. Figure 1 depicts the network architectures mentioned above, while Figure 2 illustrates the block architecture of a multilayer 2-degree node as considered in this study. More specifically, a router contains multiple line card chassis, and each one contains up to 16 line cards. Each line card includes a variable amount of ports. The router ports are connected to a transponder through a gray interface or they directly hold a pluggable. Then the signal is fed to the add/drop layer of the OXC. The internal representation of the ROADM is taken from [7] and will later be valuable for the overall cost calculations.

### B. Network Properties

We further consider a static demand matrix  $D$  with  $D_{s,d}$  being the traffic requested from node  $s$  to node  $d$  for  $(s,d) \in U$ . Demands are being served using the lightpaths in  $P$ , where  $P_{i,j}$  with  $(i,j) \in U$  includes all candidate lightpaths connecting node  $i$  with  $j$ . Each lightpath  $p \in P$  has a particular length of  $\text{length}(p)$ , starts at a node  $\text{start}(p)$ , terminates at a node  $\text{end}(p)$ , and travels through all the in-between nodes in the optical domain. The combination of all adjacent optical channels is given by  $C$ , with  $C_x$  being all the adjacent channels that reserve  $x$  continuous slots. Assuming FlexGrid with a slot width granularity of 12.5 GHz in the C-Band, we define the enumeration  $S$ , which contains all frequency slot indices ranging from 1 to 320. For example,  $C_4$  would be the set of all combinations of four contiguous slots from 1 to 320, i.e.  $\{1, 2, 3, 4\}, \{2, 3, 4, 5\}, \{3, 4, 5, 6\}, \dots, \{317, 318, 319, 320\}$ .

An optical transmission module is capable, due to its elasticity, of a set of transmission modes  $t \in T$  being described by a tuple  $(d_t, r_t, b_t, c_t, t_t)$ , where:

- $d_t$  is the optical reach in kilometers, assuming Erbium-Doped Fiber Amplifier (EDFA) amplification in every

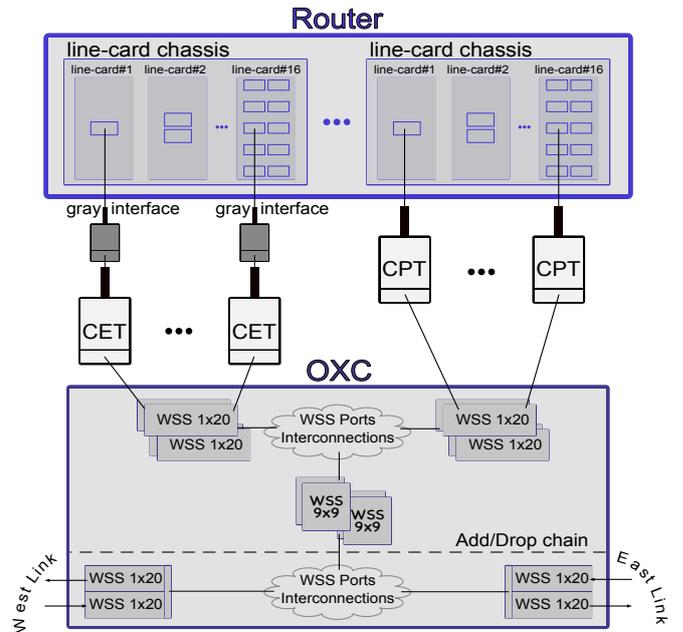


Fig. 2. The node architecture. The OXC is designed according to [7].

span of 80 km.

- $r_t$  is the data rate.
- $b_t$  is the number of frequency slots occupied as a multiple of 12.5 GHz.
- $c_t$  is the cost of the module.
- $t_t$  is the module type, either a CPT or a CET.

We narrowed these values down to Table II. Table IIa shows the properties of the OpenZR+ modules [1], while Table IIb and Table IIc show the transponder properties. The Table IIb assumes somewhat conservative values, while IIc considers more advanced technology and thus more powerful transponders with greater optical reach. This increase comes at the price of more bandwidth use. The reason for using two tables for the transponder properties is to demonstrate the influence of relatively weak versus powerful transponders as well as to demonstrate the flexibility of parametrization of the optimization ILP model. The values were derived from publications like [8], [4], from GNPY [9], and through some valuable feedback from industry partners to confirm their plausibility as products in the market. We generalized the cost of a CPT to be 40% of the CET [3], thus setting them to 8 and 20 cost units, respectively.

TABLE I  
PROPERTIES OF THE LINE CARDS.

Line cards $L$		
$n_l$ (ports)	$r_l$ port rate (Gbps)	$c_l$ (cost units)
10	100	26.72
2	400	29.36
1	1000	31.99

### C. Grooming and No-Grooming

Finally, we distinguish between two operating modes: the *grooming* and *no-grooming* modes. Electrical grooming is a

TABLE II  
PROPERTIES OF THE OPTICAL TRANSMISSION MODULES

(a) Pluggable modules

CPT		
$r_t$ (Gbps)	$d_t$ (km)	$b_t$ ( $\times 12.5$ GHz)
400	480	6
300	1600	6
200	2880	6
100	5840	4

(b) Transponder modules (conservative)

CET <i>conservative</i>		
$r_t$ (Gbps)	$d_t$ (km)	$b_t$ ( $\times 12.5$ GHz)
800	160	8
700	200	8
600	240	6
500	480	6
400	880	6
300	2080	6
200	6120	6
100	9260	4

(c) Transponder modules (advanced)

CET <i>advanced</i>		
$r_t$ (Gbps)	$d_t$ (km)	$b_t$ ( $\times 12.5$ GHz)
800	400	10
700	700	10
600	1200	8
500	2800	8
400	4400	8
300	5080	8

technique that serves to increase the spectrum utilization in the network by combining different optical signals into the same one. For this to happen, all the optical signals of interest must be converted to electrical signals and then be modulated together into the same optical channel. For example, when different lightpaths pass through the same links, network efficiency might benefit if these optical signals are grouped into a single lightpath. Such network benefits become less obvious for networks leveraging the FlexGrid or PCS technology, as each optical signal can theoretically have a precisely customized optical channel. Also, grooming techniques might increase the complexity of the path-finding algorithm involved or even yield increased hardware usage if not parametrized correctly. For these reasons, we include these two modes and measure the influence of electrical grooming in our scenarios.

## V. METHODOLOGY

Following, we will present the procedure used to solve the resource allocation problem in optical networks, considering the architecture of the previous chapter. More precisely, given a network graph  $G$ , some resources like transmission modes  $T$ , line cards  $L$ , and optical channels  $C$ , the task is to derive the optimal resource allocation for a static traffic matrix  $D$ . Optimality is measured against the costs of the resources used as presented in Section IV. This is a network dimensioning problem; hence there is no resource limitation and traffic blocking. However, no overdimensioning is performed since that would not change the results of a comparative study like

this one. The comparison will be between a HbH, a multilayer, or a hybrid network architecture and configurations using only CPTs, only CETs, or both. All these cases are treated by feeding in different inputs to an ILP, as described next.

### A. ILP

As already mentioned, the primary task of the ILP is to determine the hardware resources to serve all demand pairs in the network. Different hardware presents different properties, which the ILP needs to consider to produce a cost-efficient configuration successfully.

The ILP is inspired from [8] while also incorporating modifications to support our specific scenario. Following, we present the complete ILP formulation and highlight the main differences from its predecessor. We will show two versions, one with electrical grooming and one without. First, we will describe the ILP input, variables, and constraints for the grooming case, and later we will point out the differences for the no-grooming version.

1) *Input*: This subsection will describe the model's parameters, which are treated as constants by the ILP. We form  $B$  as the set of all spectrum slot requirements  $b_t$  for  $t \in T$ . The candidate lightpaths  $P$  are collected using the k-shortest paths between all node pairs  $(i, j) \in U$  to populate  $P_{i,j}$ .

The costs of the hardware resources used, play the role of the objective function coefficients. More specifically, we use:

- $c_t$  being the cost of a module operating at transmission mode  $t$ .
- $c_l$  being the cost of a line card  $l$ .
- $c_{lcc}$  being the cost of a line card chassis.
- $c_\phi$  being the cost of an end-to-end fiber.
- $c_{\phi km}$  being the cost of a fiber per kilometer.

Precise values of these costs are found in Section IV. We set  $c_\phi$  to 100 cost units per fiber and  $c_{\phi km}$  to 0.5 cost units per fiber kilometer.

2) *Variables*: Following, we present the variables of the ILP system:

- $f_{s,d}^{i,j} \in \mathbb{R}_{\geq 0}$  ( $(s,d), (i,j) \in U$ ) represents the traffic flow from source node  $s$  to destination node  $d$  using a lightpath from  $i$  to  $j$ .
- $x_{p,t} \in \mathbb{Z}_{\geq 0}$  ( $p \in P, t \in T$ ) is the number of deployed lightpaths  $p$  using transmission mode  $t$ . To have a valid lightpath-transmission mode pair, the optical reach of  $t$  needs to support the distance covered by the lightpath  $p$ .
- $u_{p,c} \in \mathbb{Z}_{\geq 0}$  ( $p \in P, c \in C$ ) is the number of deployed lightpaths  $p$  using channel  $c$ .
- $z_{v,l} \in \mathbb{Z}_{\geq 0}$  ( $v \in V, l \in L$ ) is the number of line cards of type  $l$  at node  $v$ . To use a transmission mode  $t$  in a line card, the port rate must be supported ( $r_l \geq r_t$ ).
- $h_v \in \mathbb{Z}_{\geq 0}$  ( $v \in V$ ) is the number of line card chassis at node  $v$ .
- $\phi_e \in \mathbb{Z}_{> 0}$  ( $e \in E$ ) is the number of fibers at edge  $e$ .

To get the number of line cards  $l$  per node, we need to create two more variables to represent an internal state:

- $\hat{z}_{v,l} \in \mathbb{Z}_{\geq 0}$  ( $v \in V, l \in L$ ) represents the line card  $l$  type ports that can be used at node  $v$  in order to serve the deployed lightpaths crossing this node.

- $\hat{z}_{v,l} \in \mathbb{Z}_{\geq 0}$  ( $v \in V, l \in L$ ) signifies the requested number of ports at node  $v$  that can be allocated either with a line card of type  $l$  or any lower rate one.

3) *Constraints and Objective Function:* Following, we present the constraints:

$$\sum_{i \in V \setminus \{v\}} f_{s,d}^{i,v} - \sum_{j \in V \setminus \{v\}} f_{s,d}^{v,j} = \begin{cases} -D_{s,d}, & \text{if } v = s \\ +D_{s,d}, & \text{if } v = d \\ 0, & \text{if } v \neq s, d \end{cases} \quad (1)$$

$$\forall (s, d) \in U, v \in V$$

$$\sum_{(s,d) \in U} f_{s,d}^{i,j} \leq \sum_{p \in P_{i,j}, t \in T} r_t \cdot x_{p,t} \quad \forall (i, j) \in U \quad (2)$$

$$\sum_{p \in P_{i,j}, t \in T | \text{length}(p) > d_t} x_{p,t} = 0 \quad \forall (i, j) \in U \quad (3)$$

$$\sum_{t \in T | b_t = b} x_{p,t} = \sum_{c \in C_b} u_{p,c} \quad \forall (p, b) \in P \times B \quad (4)$$

$$\sum_{c \in C, p \in P | s \in c \wedge e \in p} u_{p,c} \leq \phi_e \quad \forall (e, s) \in E \times S \quad (5)$$

$$\hat{z}_{v,l} = \sum_{p \in P, t \in T | (\text{start}(p) = v \vee \text{end}(p) = v) \wedge r_l \geq r_t} x_{p,t} \quad \forall (v, l) \in V \times L \quad (6)$$

$$\hat{z}_{v,l} = \begin{cases} \hat{z}_{v,l} - \hat{z}_{v, l_{\text{low}}}, & \text{if } l_{\text{low}} = \text{lower}(L, l) \\ \hat{z}_{v,l}, & \text{if } \nexists \text{lower}(L, l) \end{cases} \quad (7)$$

$$\forall (v, l) \in V \times L$$

$$\sum_{l \in L | r_l \geq r_{\hat{l}}} z_{v,l} \cdot n_l \geq \sum_{l \in L | r_l \geq r_{\hat{l}}} \hat{z}_{v,l} \quad \forall (v, \hat{l}) \in V \times L \quad (8)$$

$$h_v \geq \sum_{l \in L} z_{v,l} / N_{\text{fcc}} \quad \forall v \in V \quad (9)$$

Constraint (1) holds the flow constraints. Each node can be either a traffic generator ( $v = s$ ), a traffic sink ( $v = d$ ), or a mediator, which will forward the traffic ( $v \neq s, d$ ). The demands are being routed because of (2). This inequality obliges every flow to be served using a combination of a lightpath and a transmission mode. Equality (3) excludes such invalid combinations because of optical reach limitations. Later in (4), the specific super-channels are being chosen. The equation serves both the spectrum contiguity and continuity constraints. Inequality (5) handles the allocation of optical fibers. If the same super-channel is allocated more than once along a lightpath  $p$ , then more fibers need to be used. Equalities (6) and (7) constitute an internal state to precisely calculate the number of line cards needed. (8) takes over the final allocation of ports in terms of line cards. (9) reserves the line card chassis needed by accumulating the line cards in groups of  $N_{\text{fcc}}$ .

The objective function to be minimized is

$$\sum_{p \in P, t \in T} c_t \cdot x_{p,t} + \sum_{v \in V, l \in L} c_l \cdot z_{v,l} + \sum_{v \in V} c_{\text{fcc}} \cdot h_v + \sum_{e \in E} (c_\phi + \text{length}(e) \cdot c_{\phi \text{km}}) \cdot \phi_e. \quad (10)$$

The optimization target solely depends on the hardware equipment cost and not on network efficiency metrics. However, this ILP model implicitly works towards improving such network metrics since better network metrics lead to less equipment needed and vice versa. We will introduce some of these network metrics in Section VII.

4) *Grooming restricted version:* For the ILP version where grooming is not allowed, it is enough to substitute the variable  $x_{pt}$  with the following

- $x_{s,d}^{p,t} \in \mathbb{Z}_{\geq 0}$  ( $p \in P, t \in T$ ) is the number of deployed lightpaths  $p$  using transmission mode  $t$  for the flow between node  $s$  and node  $d$ .

Subsequently, we will need to substitute all constraints involving  $x_{pt}$  with the expression  $\sum_{(s,d) \in U} x_{p,t}^{s,d}$  except for constraint number (2), for which we must not in order to keep track of single end-to-end flow allocations. Thus, constraint (2) becomes

$$f_{s,d}^{i,j} \leq \sum_{p \in P_{i,j}, t \in T} r_t \cdot x_{p,t}^{i,j} \quad \forall (i, j) \in U, (s, d) \in U \quad (11)$$

These substitutions lead to a significant increase in the number of constraints, thus requiring more memory resources.

## VI. SCENARIO

This chapter will present a concrete description of our use cases. We also introduce the specific parameters of our model. Finally, we will go through the simulation setup and some implementation details.

### A. Network Parameters

As described in Section IV, there are several input parameters to be decided. Some of them are fixed, while others vary to observe their influence.

1) *Fixed Input Parameters:* After a series of simulations, the below parameters were chosen as fixed for the present study. The resource costs are kept constant to the values introduced in Section IV. The k-shortest path to calculate the set of candidate lightpaths  $P$  is fixed to  $k = 2$ . Each link has 320 available frequency slots. Finally, the static traffic matrix is left unchanged throughout all the simulations. The matrix  $D$  was artificially generated using a deterministic negative exponential relationship with respect to the node pair distance. Figure 3 shows the demand matrix on the left blue axis as considered for the US topology with a maximum demand value of 900 Gbps, a minimum of 50 Gbps, and the red dashed line shows the average demand value across all nodes pairs to 130 Gbps. The reader can see the node pair distances on the right orange axis.

2) *Varying Input Parameters:* To showcase the impact of some of the model inputs, we carried out a parametric sweep across all the combinations of the following parameters:

- *TopologyType* is either a US or a German topology. The two networks are shown in Figure 4. The topologies and node coordinates were retrieved from [10].
- *OXCPercentage* (OXCP) incorporates the percentage of OXCs in the network. The parameter values are increasing at a granularity of 10%, i.e. 0%, 10%, 20%, ..., 100%. The available OXCs are positioned in the network with

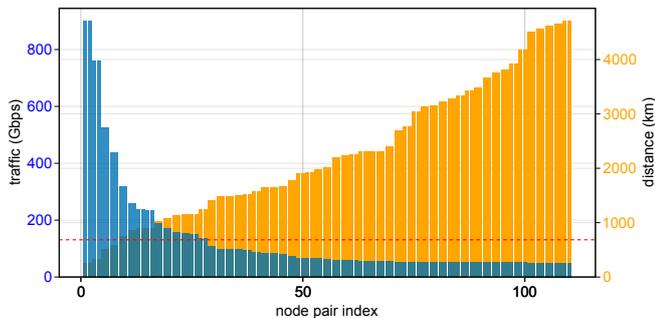


Fig. 3. The demand matrix  $D$  for the US network.

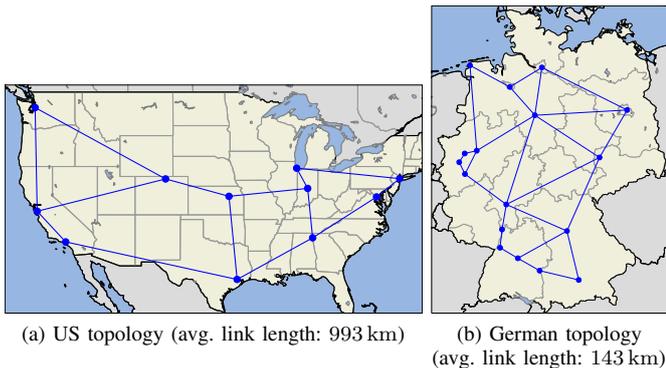


Fig. 4. The network topologies.

decreasing order at the nodes with higher betweenness

- *ILPType* (ILPT) is the two operating modes and can be Groom or NoGroom.
- *TransType* (TransT) is the type of transmission modules used. Although the transmission properties tuples always remain the same as in Table II, we vary this parameter as follows:
  - *PurePluggables* (PP) contains only Table IIa.
  - *PureTransponders* (PT) contains only Table IIb.
  - *PureTranspondersBoosted* (PTB) contains Table IIb and Table IIc.
  - *Mixed* (M) contains Table IIa and Table IIb.
  - *MixedBoosted* (MB) contains all tuples in Table II.

The results of the parametric study are in Section VII.

### B. Simulation Setup

The simulation, meta-analysis, and data visualization [11] were done using software written in Julia [12]. In particular, we used the JuMP framework [13] to model and formulate the ILP. The mathematical programming solver IBM CPLEX version 12.10 was used as a backend. We conducted our parameter study averaging over 4 seeds. For every seed, 8 threads were allocated leveraging the opportunistic mode of CPLEX, which enables the solver to aggressively search for the best solutions without caring for reproducibility. The relative tolerance on the ILP gap was set to 4%, and the time limit to stop any execution at 8 hours.

## VII. RESULTS

To interpret the impact of the variable parameters on network efficiency, we need to use some network metrics. Similar to [6], [5], we define the following:

- *Port usage* is the absolute number of router ports used.
- *Spectrum usage* is the total normalized spectrum allocated throughout the whole network.
- *Regenerations* are the total absolute amount of times a signal regenerates.
- *Latency* is the total normalized network latency as a summation of all end-to-end traffic requests. We assume the speed of light in the fiber to be 199 861.213 km/sec.
- *Power consumption* is the total normalized power consumption of the network. We consider in line with [6], [14] CPTs of 15 Watt, CETs of 150 Watt, router ports of 36 Watt, and gray transceivers of 8 Watt.
- *Hardware Costs* are the total CAPEX costs of the network, as explained in Section IV. For the OXC, we consider a CDC, FlexGrid ROADM with the architecture of Figure 2, and the costs are calculated using the assumptions of [7].

The normalization happens with respect to the maximum value in each case, as shown later.

### A. Port Usage

Figure 5 shows the port usage for both the US and the German topology. The available ports are the number of line cards times their port density, and the used ports are the allocated ports from the solver. The  $x$ -axis on the plots signifies the parameter in use, and the  $y$ -axis is the average of all simulations holding this parameter value steady. For example, we see that having 0% OXCs yields the highest port usage for the US topology on average. The ILP tries to keep the used ports close to the available number in order not to pay for an extra line card.

In the same figure, we witness a behavior that will persist throughout most results. Namely, the network metric is rather bad when on 0% OXCs and exponentially improves as we increase the OXC participation. However, the improvement rate decreases as the OXC percentage grows, and the metric converges to a particular number with negligible differences between the other high OXC values. We call it the *sufficient convergence point* where the knee of the curve is, and more OXCs will hardly make a difference (e.g., 50% for the US and 30% for Germany). So, unfortunately, there is not a panacea OXC percentage value for all networks, and finding this point depends very much on the topology in use. For example, the German topology is more interconnected than the US; thus, having fewer OXCs is fine.

Regarding the *TransType* parameter and due to the short distances in the German topology, we can observe that there is substantially no difference between using CPTs or CETs for the port usage metric. This is not the case for the US topology, where stronger transmission technology is needed to cross big distances. For this reason, the mixed Boosted case performs better than using only CPTs.

Also, it becomes obvious that no-grooming operation is generally much more resource-hungry than the grooming case.

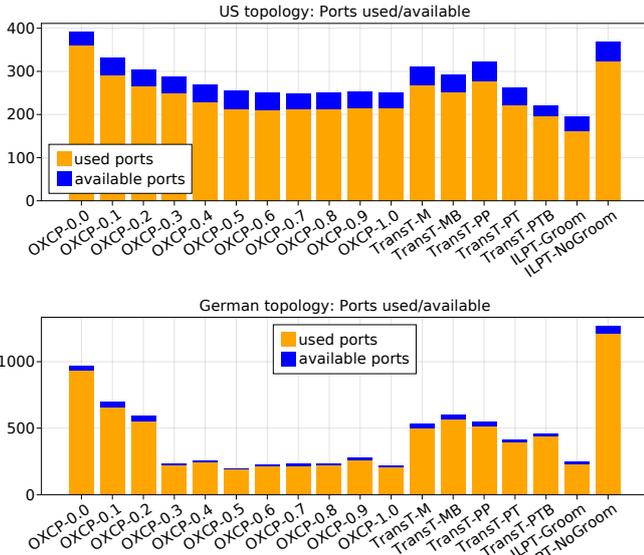


Fig. 5. Port usage parameter sweep.

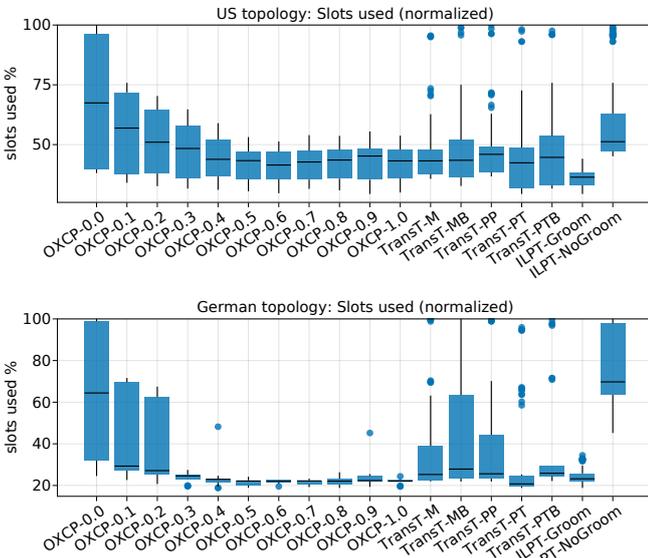


Fig. 6. Slot usage parameter sweep.

### B. Spectrum Usage

The spectrum allocation follows the same pattern as the port usage. This time, the box plots are presented in Figure 6. The crossbars span the Interquartile Range (IQR), with the midline marking the median, and the error bar whiskers span up to  $1.5 \times \text{IQR}$ . Again, we witness that having 0% OXCs yields the highest spectrum allocation. Hence, all other values will be normalized based on this maximum.

The slight increase of the powerful transponders (i.e. Boosted) compared to conventional transponders happens due to the increased problem complexity. The *MixedBoosted* and *PureTranspondersBoosted* are supersets of *Mixed* and *PureTransponders*, respectively, and naturally, they can only do better. The problem is that the more transponder tuples the

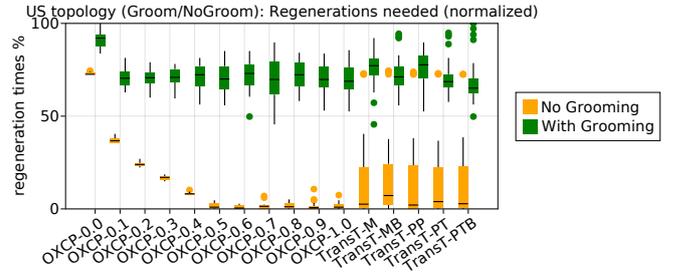


Fig. 7. Regeneration times parameter sweep.

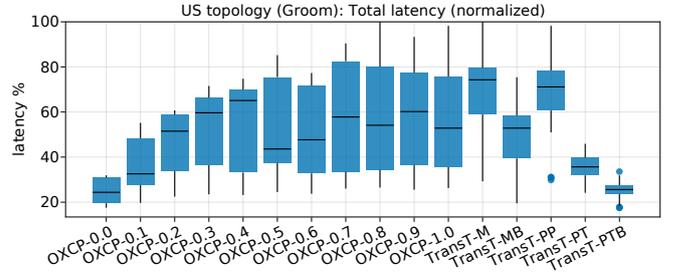


Fig. 8. Latency parameter sweep.

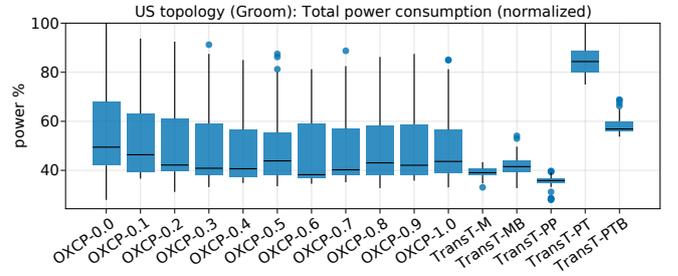


Fig. 9. Regeneration times parameter sweep.

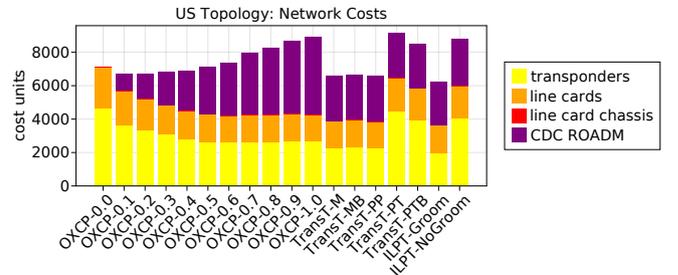


Fig. 10. Hardware costs parameter sweep.

ILP considers, the harder the problem becomes and the more variance in the results. The boxplots detect this uncertainty with a rather long crossbar and even longer whiskers. However, much of this variance comes from the no-grooming case due to its extreme behavior. This can be seen better in the next subchapter and Figure 7.

### C. Regenerations

We will now continue only with the results of the US topology since the behavior of the two networks is similar. Observing Figure 7, we indeed conclude that the no-grooming

case introduces some more variance in our results for the different transmission modules. In addition, when grooming is not allowed, regeneration times are better minimized, approaching values close to zero. However, more OXCs are needed to reach the sufficient convergence point compared to the grooming case (50 % against 10 % OXCs).

#### D. Latency

Another interesting impact is that lower latencies are achieved by decreasing the OXCs or preferring CETs over CPTs. Figure 8 shows exactly this for the US topology and grooming allowed. The major factor of these latencies is the propagation delay, as is commonly the case in core networks. When the OXC percentage is low, the ILP prefers to take the shortest path to route all demands. As the OXC percentage increases, more opportunities emerge to optimize the resource allocation by grooming and optically bypassing the signals using the second shortest paths, thus increasing the end-to-end delay. On the other hand, powerful transponders allow the solver to not commit to nearby regeneration points but route the traffic directly through the shortest path.

#### E. Power Consumption

Power consumption is a part of the network operator's OpEx costs. Figure 9 shows that a hybrid architecture consisting of both CPTs and CETs can have a power consumption very close to the one using only CPTs. We observe that the pure CET case is, in contrast, very power-hungry.

#### F. Hardware Costs

Figure 10 shows the overall network node costs of the US topology for the different parameters. The line card chassis are mostly constant to 1, so no multi-chassis routers are really needed with this traffic matrix. As the OXC percentage and the costs for ROADMs increase, the need for transponders decreases. Interestingly, overall costs balance around the same value 6500, up to 50 % OXCs. Regarding the transmission modules, once again, the mixed case is standing very close to the pure CPT case, which appears much more economical than the pure CET cases.

### VIII. CONCLUSION

Overall in this paper, we discussed the problem of choosing between different core network architectures. We considered a multilayer architecture versus a HbH and networks using only CETs modules versus only CPTs. We showcased that the optimal behavior comes with a hybrid architecture for both cases, where we can have the best of two worlds. Unfortunately, finding the best architecture recipe is very much dependent on the network topology and the underlying resources, which were treated as problem parameters. We provided an ILP formulation to find this optimal point and examined it for several problem instances. We found that overly increasing the OXC coverage might saturate the network metrics, where no further win will be possible. This had more impact on a higher interconnected topology like the German network. Regarding the transmission

modules, using both CETs and CPTs demonstrates costs close to the pure CPT case while at the same time approaching the better network metrics of the CETs. This was more apparent for large distance networks like the US backbone. The problem instance fed to the ILP can easily be adapted to any network operators' present or future perspectives.

### REFERENCES

- [1] "White paper: OpenZR+ 400G Digital Coherent Optics for Multi-Haul," OpenZR+ Multi-Source Agreement, Tech. Rep., September 2020, Accessed on 28.03.2022. [Online]. Available: [https://openzrplus.org/site/assets/files/1074/openzrplus\\_whitepaper\\_-\\_sept\\_29\\_2020\\_final.pdf](https://openzrplus.org/site/assets/files/1074/openzrplus_whitepaper_-_sept_29_2020_final.pdf)
- [2] Colt, "Colt takes network innovation to new heights with a 400G-capable Routed Optical Networking solution on its IQ Network," <https://www.colt.net/resources/colt-takes-network-innovation-to-new-heights-with-a-400g-capable-routed-optical-networking-solution-on-its-iq-network/>, Accessed on 28.03.2022.
- [3] P. Wright, R. Davey, and A. Lord, "Cost Model Comparison of ZR/ZR+ Modules Against Traditional WDM Transponders for 400G IP/WDM Core Networks," in *2020 European Conference on Optical Communications (ECOC)*, 2020.
- [4] A. Eira and J. Pedro, "On the Comparative Efficiency of Next-Generation Coherent Interfaces for Survivable Network Design," in *17th International Conference on the Design of Reliable Communication Networks (DRCN)*, 2021.
- [5] S. Melle, T. Zami, O. Bertran-Pardo, and B. Lavigne, "Comparing IP-over-WDM Architectures & WDM Transport Technologies in Metro, Regional and Long-Haul Networks," in *2021 Optical Fiber Communications Conference and Exhibition (OFC)*, 2021.
- [6] A. Gumaste, M. Sosa, H. Bock, and P. Kandappan, "Optimized IP-over-WDM core networks using ZR+ and flexible muxponders for 400 Gb/s and beyond," *Journal of Optical Communications and Networking*, vol. 14, no. 3, pp. 127–139, 2022.
- [7] F. Rambach, B. Konrad, L. Dembeck, U. Gebhard, M. Gunkel, M. Quagliotti, L. Serra, and V. López, "A Multilayer Cost Model for Metro/Core Networks," *Journal of Optical Communications and Networking*, vol. 5, no. 3, pp. 210–225, 2013.
- [8] P. Papanikolaou, K. Christodouloupoloulos, and E. Varvarigos, "Multilayer flex-grid network planning," in *2015 International Conference on Optical Network Design and Modeling (ONDM)*, 2015, pp. 151–156.
- [9] A. Ferrari, M. Filer, K. Balasubramanian, Y. Yin, E. Le Rouzic, J. Kundrát, G. Grammel, G. Galimberti, and V. Curri, "GNPy: an open source application for physical layer aware open optical networks," *Journal of Optical Communications and Networking*, vol. 12, no. 6, pp. C31–C40, 2020.
- [10] S. Orłowski, M. Pióro, A. Tomaszewski, and R. Wessäly, "SNDlib 1.0—Survivable Network Design Library," in *3rd International Network Optimization Conference (INOC)*, 2007, <http://sndlib.zib.de>, Accessed on 28.03.2022.
- [11] S. Danisch and J. Krumbiegel, "Makie.jl: Flexible high-performance data visualization for Julia," *Journal of Open Source Software*, vol. 6, no. 65, 2021.
- [12] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, "Julia: A Fresh Approach to Numerical Computing," *SIAM review*, vol. 59, no. 1, pp. 65–98, 2017.
- [13] I. Dunning, J. Huchette, and M. Lubin, "JuMP: A Modeling Language for Mathematical Optimization," *SIAM Review*, vol. 59, no. 2, pp. 295–320, 2017.
- [14] W. Van Heddeghem and F. Idzikowski, "Equipment power consumption in optical multilayer networks – source data," Tech. Rep., January 2012.

# Control-Plane Traffic Modelling for Connection Management in T-SDN Optical Networks using Transport PCE and OpenROADM

Shabnam Sultana<sup>1,2</sup>, Ronald Romero Reyes<sup>2</sup>, Thomas Bauschert<sup>2</sup>

<sup>1</sup> Highstreet Technologies GmbH, 12159 Berlin, Germany

Email: shabnam.sultana@highstreet-technologies.com

<sup>2</sup> Chair of Communication Networks

Technical University of Chemnitz, D - 09107 Chemnitz, Germany

**Abstract**—A crucial component of the architecture of software defined networks is the control plane. With the increasing traffic demand and traffic dynamics, the scalability and resiliency of the control plane infrastructure gains importance. While there is a plethora of studies being performed on control plane optimization for packet-switched networks, these issues are yet to be thoroughly explored for software defined transport networks (T-SDN). Control plane optimization involves not only the placement of controllers, but also the appropriate design of the signalling network (the so-called DCN), which requires knowledge of the control traffic demand. In this paper, considering a T-SDN based network scenario comprising Transport PCE controllers and OpenROADMs, we focus on the modelling and evaluation of the control traffic that stems from the signalling procedures for lightpath setup and teardown. Our modelling framework unveils the control traffic dependency on the number of transponders and Reconfigurable Optical Add/Drop Multiplexers (ROADMs) involved in the setup and teardown of optical connections. Furthermore, the proposed framework paves the way for the design of resilient DCNs for disaggregated T-SDN enabled optical networks.

**Index Terms**—Software defined networking, optical networks, traffic modelling, OpenROADM, Transport PCE.

## I. INTRODUCTION

As software-defined networking (SDN) is considered the most promising approach for next-generation networks, aspects like scalability, resiliency, and reactivity of the SDN control-plane have gained momentum. Consequently, numerous studies have come up investigating these aspects and proposing novel approaches for optimizing the SDN control-plane. The survey in [1] has reviewed 99 such published studies, proposing novel approaches and utilizing a wide range of algorithms and heuristics with diverse optimization objectives such as optimizing the number and placement of controllers. With the increase in traffic volume and dynamics, the analysis of the SDN control-plane performance and its optimization will become even more crucial. So far only a few studies focused on the optimal controller placement that minimizes the control or signalling traffic, i.e. the traffic

between the control plane and the data plane and between the controllers themselves. In fact, comprehending and modelling the control traffic could help determining the optimal design of the control-plane not only in terms of the number and location of controllers but also in terms of computing resource requirements. While there have been some studies focusing on the control traffic in SDN-based packet-switched networks, as per our knowledge, this paper is the first one which presents a control traffic model for software defined optical transport networks (T-SDN).

T-SDN proposes a control-plane architecture for the control and management of transport networks that may involve multi-layer, multi-domain and multi-vendor scenarios, with the potential of providing advantages for network operators such as cost reduction and profit maximization [2]. Besides, consortia like OpenConfig and OpenROADM address the heterogeneous nature and complexity of optical equipment by defining vendor-agnostic standard YANG models that can be used by the Network Configuration (NETCONF) protocol. Furthermore, the OpenROADM Multi-Source Agreement (MSA) proposes a complete disaggregation of the network by modelling the interoperability capabilities of Reconfigurable Optical Add/Drop Multiplexers (ROADM) along with transponders and pluggable optics [3]. However, unlike packet-switched SDN networks, in which flow configuration procedures are well standardized, in disaggregated T-SDN optical networks the sequence of controller-device operations is strongly dependent on the controller implementation. OpenROADM MSA standardizes data models that describe the devices, services and topology, but does not specify any requirements, implementation and operational aspects for OpenROADM controllers [3]. Nevertheless, the controller has to comply with the OpenROADM device models in order to communicate with the devices for read-write NETCONF operations, implying that two different controllers (if OpenROADM compliant) send and receive identical (standardized) messages to and from the devices.

The remaining sections of this paper are organized as follows. Section II provides a brief overview of previous related work. Section III contains an analysis of T-SDN

This work has been funded by the German BMBF OptiCON project and the Celtic-Plus project AI-NET-PROTECT, funded by the German BMBF (ID 16KIS1285).

control plane message flows during lightpath establishment and termination. In Section IV the data rate induced by control plane traffic is evaluated considering dynamic lightpath setups and terminations. Finally, section V concludes the paper.

## II. RELATED WORK

Concerning control-plane optimization in packet-switched networks, only a few studies have analysed the control traffic and endeavoured to determine the placement and the number of controllers that minimize the control traffic. The work in [4] models the controller placement problem using Integer Quadratic Programming with the objective of minimizing the overall capacity required for the control traffic. This study was performed on a testbed that consists of 50 Open vSwitch (OvS) NITOS nodes controlled by Kandoo controllers. An extension to this study with OpenDaylight (ODL) controllers is presented in [5]. However, both studies only model the relative increase of the control traffic as function of the number of controllers. A control traffic model capable of predicting the number of transactions/messages between switches and controllers during a connection request is presented in [6]. By evaluating the exact number of OpenFlow (OF) messages required to install traffic flows, the authors compute the control traffic in reactive OF operational-mode for a realistic topology of a nationwide ISP. This model enables the assessment of the scalability of the reactive operational-mode in practical deployment scenarios.

In optical networks, lightpath establishment and termination are key control-plane procedures for connection management that strongly impact the network performance. A full understanding of the control traffic generated by these procedures not only should help optimize the number or placement of controllers but also the design of the signalling or Data Communication Network (DCN) itself. Therefore, in this paper, we assess the control traffic by analyzing the NETCONF messages which are exchanged during the lightpath setup and termination procedures. This study is motivated by the increasing interest in the cost-efficient design of disaggregated T-SDN optical networks. In this direction, ongoing research has provided key developments that include SDN control frameworks and industry-standard implementations based on TAPI interfaces and OpenROADM device models [7]. Furthermore, in the course of disaggregated T-SDN optical networks, the work in [8] demonstrated the ability of Transport PCE controllers to automatically create and delete services in multi-vendor networks that implement OpenROADM data models. Built on the ODL controller, Transport PCE follows the OpenROADM service, network and device data models, thereby ensuring interoperability [9].

## III. ANALYSIS OF T-SDN CONTROL PLANE MESSAGES

### A. Simulation Setup and General Observations

To assess the control traffic load that stems from the lightpath setup and termination procedures, we created a simulation testbed where Transport PCE is used as control framework with OpenROADM device data models as specified by the Transport PCE project [9]. The testbed is implemented with

the Network Topology Simulator (NTS) [10], which provides a framework for the simulation of optical devices controlled through NETCONF/YANG interfaces. Valid YANG models are configured into these devices and exposed to the controller via NETCONF/YANG. The NTS testbed is used to measure the control traffic that stems from the setup and termination of lightpaths (with bit-rate 100 Gb/s) in a linear network made up of a chain of ROADMs. The lightpaths are established between the ROADMs at the network edges. The optical network is assumed to be transparent, i.e. regeneration is not required. Each lightpath is provisioned with a pair of transponders, one installed at each edge ROADM. The lightpath setup and termination procedures involve a sequence of read-write NETCONF operations performed between the controller and the devices (i.e. transponders and ROADMs). These operations are realized through the exchange of *request* and *reply* remote procedure call (RPC) signalling messages.

The simulation results evince that during lightpath setup and termination, the controller applies standard read-write operations that are used to configure network devices. These operations are defined by the exchange of messages shown in Fig 1. As can be seen, a standard read-write operation consists of  $\langle get-config \rangle$  and  $\langle edit-config \rangle$  NETCONF messages that read and edit device configuration data, respectively. These operations are performed on device *candidate* and *running* datastores. The former stores device configuration prior to committing it to the running datastore via a  $\langle commit \rangle$  operation. The latter holds the complete configuration currently active on the device. To ensure secure and reliable device configuration, the  $\langle get-config \rangle$  and  $\langle edit-config \rangle$  operations are executed with auxiliary  $\langle lock \rangle$ ,  $\langle commit \rangle$  and  $\langle unlock \rangle$  operations. Figure 1 shows that, before performing a read-write operation, the first RPC message exchanged is  $M_{lock}$ , which corresponds to the  $\langle lock \rangle$  operation whereby the controller locks the candidate and running datastores of the device. This procedure prevents overwriting interactions from other NETCONF and non-NETCONF clients (e.g. SNMP and command-line interface scripts). The device replies back with the  $R_{lock}$  message, acknowledging that the lock has been successful. At this stage, the controller is ready to perform  $\langle get-config \rangle$ ,

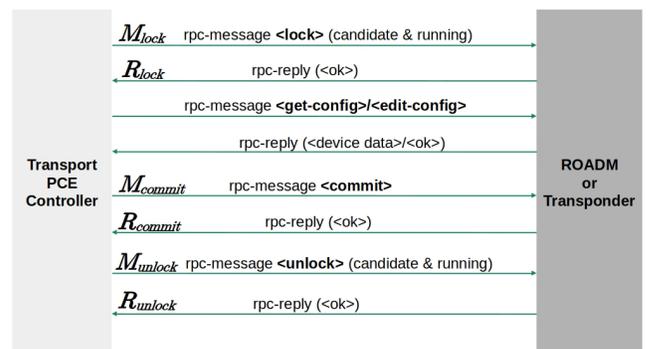


Fig. 1. Signalling messages involved in a standard read-write operation.

$\langle edit-config \rangle$  and  $\langle commit \rangle$  operations on the device configuration data. Upon completion of these operations, the  $\langle unlock \rangle$  operation is executed to unlock the datastores. Since the  $\langle lock \rangle$ ,  $\langle commit \rangle$  and  $\langle unlock \rangle$  operations are involved in every read-write operation of the lightpath setup and termination procedures, we denote these operations as transaction envelope. The total size of the controller-to-device signalling messages exchanged in this transaction envelope is given by:

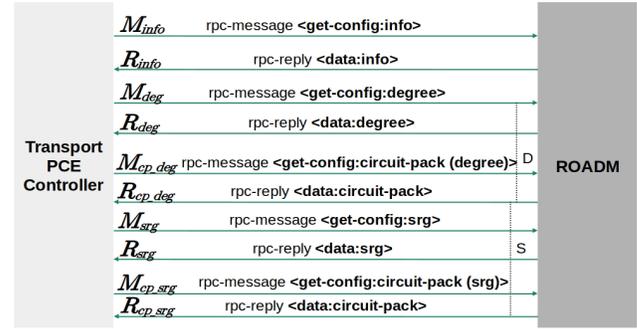
$$S(M_{TEnv}) = S(M_{lock}) + S(M_{commit}) + S(M_{unlock}) \quad (1)$$

Table I lists all NETCONF control request and reply messages observed in the testbed. The RPC requests are denoted as  $M_X$ , where  $X$  is the name of the respective operation. The corresponding RPC replies are denoted as  $R_X$ . Furthermore, for each operation  $X$ , the size of the XML content - in Kilobytes (KB) - of the request and reply messages measured in the testbed are denoted as  $S(M_X)$  and  $S(R_X)$ , respectively. As each RPC request is acknowledged by a reply, the number of RPC replies for any operation is equal to the number of the corresponding requests. This does not imply that the message sizes  $S(M_X)$  and  $S(R_X)$  are also identical, as seen in Table I. Thus, the exchange of requests and replies results in bidirectional traffic flows with asymmetric size between the controller and the devices. In the following, for ease of reading, we focus on the derivation of the equations that describe the size of the controller-to-device traffic. However, the device-to-controller traffic can easily be derived by just replacing in the equations the request message size  $S(M_X)$  by the corresponding reply message size  $S(R_X)$ , for each operation  $X$ .

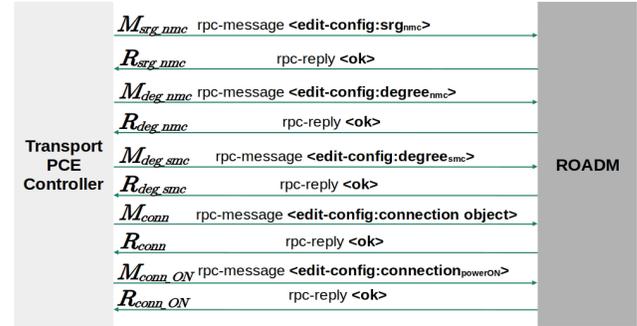
### B. Control Traffic in the Lightpath Setup Procedure

When a lightpath is set up, the controller performs NETCONF operations to configure the transponders and ROADMs to be used by the lightpath. In the following, the controller-to-device traffic that originates from these operations is described based on the observations from the simulation testbed.

1) *Controller-ROADM Traffic:* To establish a lightpath, the controller needs complete information about the current configuration of each ROADM to be traversed by the lightpath. For those ROADMs for which such information is not known yet, the controller performs the sequence of  $\langle get-config \rangle$  operations shown in Fig. 2a. By these operations, the controller inquires ROADM-related data such as the model type, the serial number, the degrees and the shared risk groups (SRG) associated to each degree. Both the degrees as well as the SRGs are composed of circuit packs. A circuit pack of a SRG consists of hardware functional blocks such as common cards and/or pluggable optics whereas a circuit pack of a degree contains wavelength selective switches, in-line amplifiers and any further equipment associated with the degree. Fig. 3 shows an example configuration of an OpenROADM compliant optical network. It contains two ROADMs, where the add/drop modules of each ROADM are represented by the SRGs. Note that in Fig. 3 three link types are depicted, namely, external links which are used to connect the ROADM to external



(a) Sequence of  $\langle get-config \rangle$  operations.



(b) Sequence of  $\langle edit-config \rangle$  operations.

Fig. 2. NETCONF operations: controller-ROADM.

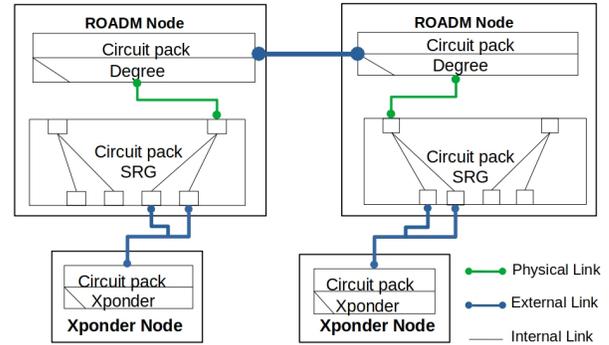


Fig. 3. Configuration of an OpenROADM compliant optical network (Example).

devices (such as transponders or other ROADMs), physical links that - within a ROADM - interconnect degrees and SRGs, and internal links that represent the interconnections between the respective circuit packs within the degree or SRG. In the lightpath setup procedure, the messages shown in Fig. 2a are required to inform the controller about the current configuration of each ROADM, especially the hardware interconnections represented by the internal links. Note that physical links are configured for each lightpath traversing a ROADM during the lightpath setup procedure. For a ROADM with a number  $D$  of degrees and a number  $S$  of SRGs, the total size  $S(M_{total_{gc}})$  of the controller-to-ROADM request

TABLE I  
NETCONF OPERATIONS PERFORMED BETWEEN THE CONTROLLER AND OPENROADM DEVICES.

Control Message Name	Description	$S(M_X)$ (KB)**	$S(R_X)$ (KB)**
$M_{lock}, R_{lock}$	RPC request and reply for the NETCONF <i>&lt;lock&gt;</i> operation.	0.64	0.40
$M_{unlock}, R_{unlock}$	RPC request and reply for the NETCONF <i>&lt;unlock&gt;</i> operation.	0.65	0.40
$M_{commit}, R_{commit}$	RPC request and reply for the <i>&lt;commit&gt;</i> operation.	0.18	0.20
$M_{info}, R_{info}$	RPC request and reply to obtain the “info” subsection from the connected ROADMs and transponders.	0.34	1.67
$M_{device}, R_{device}$	RPC request and reply to obtain the entire device data model from connected optical equipment.	0.28 <sup>X</sup> 0.56 <sup>R</sup>	15.63 <sup>X</sup> 67.77 <sup>R</sup>
$M_{deg}, R_{deg}$	RPC request and reply to obtain the “degree” information from a ROADM that comprises of the list of the supported circuit packs.	1.25	2.74
$M_{cp\_deg}, R_{cp\_deg}$	RPC request and reply enquiring the “circuit pack” model information corresponding to each degree in a ROADM.	2.08	5.33
$M_{srg}, R_{srg}$	RPC request and reply enquiring the “shared-risk group” information based on the number of SRG obtained from the info subsection in a ROADM.	1.30	2.14
$M_{cp\_srg}, R_{cp\_srg}$	RPC request and reply enquiring the “add-drop circuit packs” corresponding to each SRG in a ROADM.	1.76	10.47
$M_{srg\_nmc}, R_{srg\_nmc}$	RPC request and reply corresponding to the creation of network media channel (NMC) interfaces at the respective SRG circuit pack inside a ROADM.	2.47	0.20
$M_{deg\_nmc}, R_{deg\_nmc}$	RPC request and reply corresponding to creation of NMC interfaces at the respective degree circuit pack inside a ROADM.	2.66	0.21
$M_{deg\_smc}, R_{deg\_smc}$	RPC request and reply to create the supporting media channel (MC) at the respective degree circuit pack inside a ROADM.	2.61	0.20
$M_{conn.}, R_{conn.}$	RPC request and reply resulting in the creation of the connection object.	1.88	0.21
$M_{conn\_ON}, R_{conn\_ON}$	RPC request and reply for manipulating the power value to <i>ON</i> state in the previously created connection object resulting in service activation or lightpath establishment.	2.96	0.30
$M_{get\_conn.}, R_{get\_conn.}$	RPC request and reply enquiring the connection object as part of validating the lightpath establishment from the respective ROADM.	1.10	1.47
$M_{och\_net}, R_{och\_net}$	RPC request and reply for creating the OCH creation on the network port of the transponder in case it does not exist on the network port yet. Output power is also configured on the OCH interface.	2.75	0.20
$M_{otu\_net}, R_{otu\_net}$	RPC request and reply to create the specific OTU interface (for e.g. OTU4 for a 100 GE signal) on the supporting OCH interface on a transponder.	1.55	0.10
$M_{odu\_net}, R_{odu\_net}$	RPC request and reply to create the specific ODU interface on the corresponding OTU interface on the transponder.	1.52	0.11
$M_{eth\_client}, R_{eth\_client}$	RPC request and reply to create the corresponding Ethernet interface on the client port of the transponder, if no corresponding interface exists on the transponder client port.	1.19	0.10
$M_{gc\_c-pm}, R_{gc\_c-pm}$	RPC request and reply to verify the connection between client and network transponder circuit packs and ports by requesting for current-pm data with <i>pmParameterName/type = opticalPowerOutput</i> .	3.84	21.12
$M_{conn\_min}, R_{conn\_min}$	RPC request and reply to set the power level to a minimum value during a lightpath termination in a ROADM.	2.95	0.30
$M_{get\_ifc}, R_{get\_ifc}$	RPC request and reply to obtain the current state of the logical interfaces that were created in a device during a particular lightpath establishment.	0.10 <sup>X</sup> 2.59 <sup>R</sup>	1.10 <sup>X</sup> 3.27 <sup>R</sup>
$M_{ifc\_oos}, R_{ifc\_oos}$	RPC request and reply to update the admin state of all logical interfaces supporting the connection to <i>out-of-service</i> state. For a ROADM, they are the NMC and MC interfaces; for a transponder, they are the ODU, OTU and Ethernet interfaces.	2.02 <sup>X</sup> 2.68 <sup>R</sup>	0.10 <sup>X</sup> 0.10 <sup>R</sup>
$M_{conn\_del}, R_{conn\_del}$	RPC request and reply to delete the connection object in a ROADM.	1.32	0.10
$M_{ifc\_del}, R_{ifc\_del}$	RPC request and reply to delete the respective logical interfaces from the device that are associated with the lightpath during its termination.	0.73 <sup>X</sup> 1.84 <sup>R</sup>	0.10 <sup>X</sup> 0.10 <sup>R</sup>
$M_{get\_och}, R_{get\_och}$	RPC request and reply to fetch the respective OCH interface from a transponder.	0.25	1.15
$M_{och\_min}, R_{och\_min}$	RPC request and reply to set minimum transmit power of the OCH interface in a transponder.	1.36	0.10
$M_{get\_cp}, R_{get\_cp}$	RPC request and reply fetching the circuit pack supporting the OCH interface in a transponder.	0.52	1.60
$M_{cp\_nra}, R_{cp\_nra}$	RPC request and reply to set equipment state of respective circuit pack to <i>not-reserved-available</i> .	1.49	0.10
$M_{och\_del}, R_{och\_del}$	RPC request and reply to delete the OCH interface from a transponder.	0.61	0.10

\*\*KB = Kilobytes

<sup>X</sup>size of the corresponding request and reply messages exchanged between the controller and a transponder.

<sup>R</sup>size of the corresponding request and reply messages exchanged between the controller and a ROADM.

messages exchanged during the  $\langle get\text{-}config \rangle$  operations is calculated as:

$$\begin{aligned} S(M_{total_{gc}}) &= S(M_{info}) + S(M_{TE_{env}}) + (S(M_{deg}) \\ &\quad + S(M_{cp_{deg}}) + 2 * S(M_{TE_{env}})) * D \\ &\quad + (S(M_{srg}) + S(M_{cp_{srg}}) \\ &\quad + 2 * S(M_{TE_{env}})) * S \end{aligned} \quad (2)$$

where the size  $S(M_{TE_{env}})$  of the transaction envelope must be included since, as explained in Section III-A, each  $\langle get\text{-}config \rangle$  and  $\langle edit\text{-}config \rangle$  request and reply is accompanied by the transaction envelope. In particular, for the controller-to-ROADM traffic, each  $\langle get\text{-}config \rangle$  operation in Fig. 2a implicitly involves the  $\langle lock \rangle$ ,  $\langle commit \rangle$  and  $\langle unlock \rangle$  operations in Fig. 1. By plugging the corresponding message sizes shown in Table I into (2),  $S(M_{total_{gc}})$  in KB, is given by:

$$S(M_{total_{gc}}) = 1.81 + 6.26 * D + 6.1 * S \quad (3)$$

As for the corresponding ROADM-to-controller traffic, by following a similar analysis based on Fig. 2a, the total size  $S(R_{total_{gc}})$  - in KB - of the ROADM-to-controller reply messages that result from the  $\langle get\text{-}config \rangle$  operations is:

$$S(R_{total_{gc}}) = 2.67 + 10.07 * D + 14.61 * S \quad (4)$$

Note that the  $\langle get\text{-}config \rangle$  operations in Fig. 2a are performed only once per ROADM. It has been verified in our simulation that, if the controller has already obtained the information from a ROADM as a result of a previous lightpath establishment, the operations in Fig. 2a are not performed once again for another lightpath.

Upon completion of the  $\langle get\text{-}config \rangle$  operations in Fig. 2a, the controller has full information about the ROADM configuration. In the next step, the controller performs the necessary ROADM configuration changes to establish the lightpath. This configuration specifies how the lightpath should be switched through the circuit packs of the respective degree and SRG and is accomplished by the configuration of logical interfaces followed by the creation of a ROADM-connection object. The configuration is directed by the controller via the sequence of  $\langle edit\text{-}config \rangle$  operations as shown in Fig. 2b. (For further details about these operations see [3] and [7].) From Fig. 2b, it follows that the total size  $S(M_{total_{ec}})$  of the controller-to-ROADM request messages exchanged during the  $\langle edit\text{-}config \rangle$  operations is:

$$\begin{aligned} S(M_{total_{ec}}) &= S(M_{srg_{nmc}}) + S(M_{deg_{nmc}}) \\ &\quad + S(M_{deg_{smc}}) + S(M_{conn.}) \\ &\quad + S(M_{conn_{ON}}) + 5 * S(M_{TE_{env}}) \end{aligned} \quad (5)$$

Based on the message sizes in Table I, it follows that in (5),  $S(M_{total_{ec}}) = 19.93$  KB. A similar analysis of the ROADM-to-controller reply messages based on Fig. 2b reveals that  $S(R_{total_{ec}}) = 6.11$  KB, which is the total size of the ROADM-to-controller reply messages transferred during the  $\langle edit\text{-}config \rangle$  operations. Finally, the controller executes on the ROADM a validation check of the newly created lightpath by querying for the info subsection. Based on our

observations, the total size  $S(M_{validate})$  of the controller-to-ROADM messages transferred during the validation operation is:

$$\begin{aligned} S(M_{validate}) &= S(M_{info}) + S(M_{device}) \\ &\quad + S(M_{get_{conn.}}) + 3 * S(M_{TE_{env}}) \end{aligned} \quad (6)$$

With the message sizes in Table I, it follows that  $S(M_{validate}) = 6.4$  KB. The corresponding ROADM-to-controller message size is  $S(R_{validate}) = 73.91$  KB.

From the previous analysis, when a lightpath is established, the full configuration of a ROADM traversed by the lightpath results in a controller-to-ROADM total message size given by:

$$\begin{aligned} S(M_{ROADM_c}) &= S(M_{total_{gc}}) + S(M_{total_{ec}}) \\ &\quad + S(M_{validate}) \end{aligned} \quad (7)$$

which can be calculated by applying (2),(5),(6). Equivalently, the ROADM-to-controller total message size is calculated as:

$$\begin{aligned} S(R_{ROADM_c}) &= S(R_{total_{gc}}) + S(R_{total_{ec}}) \\ &\quad + S(R_{validate}) \end{aligned} \quad (8)$$

When the controller has no prior information about the ROADM configuration,  $S(M_{total_{gc}})$  and  $S(R_{total_{gc}})$  are not zero as the  $\langle get\text{-}config \rangle$  operations in Fig. 2a must be applied. In this case, based on Table I, (7) and (8), the total request/reply message sizes in KB are as follows:

$$S(M_{ROADM_c}) = 28.14 + 6.26 * D + 6.1 * S \quad (9)$$

$$S(R_{ROADM_c}) = 82.69 + 10.07 * D + 14.61 * S \quad (10)$$

Otherwise, if the controller has knowledge of the configuration,  $S(M_{total_{gc}})$  and  $S(R_{total_{gc}})$  are zero. In this case the total request/reply message sizes in KB are as follows:

$$S(M_{ROADM_c}) = 28.14 \quad (11)$$

$$S(R_{ROADM_c}) = 82.69 \quad (12)$$

2) *Controller-Transponder Traffic:* Open-ROADM MSA defines a transponder as an xponder device with the same rate on the client-side and on the network side but without an internal cross-connect. The exchange of signalling messages between the controller and a transponder during the lightpath setup procedure is shown in Fig. 4. As with ROADMs, in the first step, the controller requests from each transponder assigned to the lightpath configuration data such as identification, model type, life-cycle and information about client-side and network-side circuit packs. This process involves  $\langle get\text{-}config \rangle$  operations that use the request messages  $M_{info}$ ,  $M_{device}$ , and the reply messages  $R_{info}$ ,  $R_{device}$  - see Table I. From Fig. 4, it follows that the total size  $S(M_{gc})$  of the controller-to-transponder messages transferred during these  $\langle get\text{-}config \rangle$  operations is:

$$S(M_{gc}) = S(M_{device}) + S(M_{info}) + 2 * S(M_{TE_{env}}) \quad (13)$$

Based on Table I and (13), we have that:  $S(M_{gc}) = 3.56$  KB. A similar analysis of the transponder-to-controller reply messages yields  $S(R_{gc}) = 19.30$  KB.

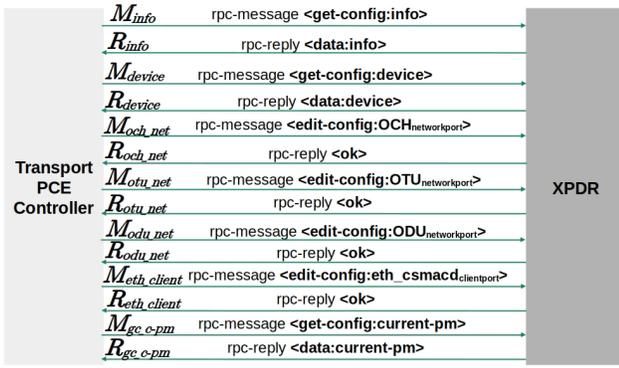


Fig. 4. NETCONF operations: controller-transponder.

In the next step, as seen in Fig. 4, the controller configures the transponders with the data required to establish the lightpath. This involves the creation of logical interfaces on the network and/or client ports of the transponders. (OpenROADM MSA exemplifies the creation of these interfaces during the provisioning of a 100GE client signal in [3].) The interfaces on the network-side ports are configured by using the messages  $M_{och\_net}$ ,  $M_{otu\_net}$  and  $M_{odu\_net}$ , which are RPC requests that create OCH, OTU4 and ODU4 interfaces, respectively. The RPC message  $M_{eth\_client}$  defines an interface of type “ethernetCsmacd” on the corresponding client port. This configuration is accomplished by the  $\langle edit-config \rangle$  operations shown in Fig. 4. Therefore, the total size  $S(M_{ec})$  of the controller-to-transponder messages transferred in the configuration procedure is:

$$S(M_{ec}) = S(M_{och\_net}) + S(M_{otu\_net}) + S(M_{odu\_net}) + S(M_{eth\_client}) + 4 * S(M_{TEnv}) \quad (14)$$

Finally, the controller affirms the transponder configuration via the message  $M_{gc\_c-pm}$  the size of which can be expressed as follows:

$$S(M_{get\_pm}) = S(M_{gc\_c-pm}) + S(M_{TEnv}) \quad (15)$$

From the previous analysis it follows that, in case of a lightpath establishment, the total size of the messages used for a full configuration of a transponder is given by:

$$S(M_{XPDRc}) = S(M_{gc}) + S(M_{ec}) + S(M_{get\_pm}) \quad (16)$$

By taking the sizes of the respective request and reply messages from Table I, we have that:  $S(M_{ec}) = 12.89$  KB. Equivalently, a similar analysis on Fig. 4 for the transponder-to-controller messages yields  $S(R_{ec}) = 4.51$  KB. Furthermore, it can be verified that:  $S(M_{get\_pm}) = 5.31$  KB and  $S(R_{get\_pm}) = 22.12$  KB. Hence, the total sizes of the controller-to-transponder and transponder-to-controller messages are  $S(M_{XPDRc}) = 21.76$  KB and  $S(R_{XPDRc}) = 45.92$  KB, respectively.

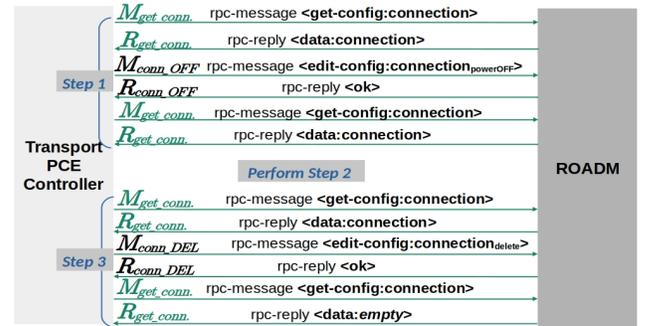
### C. Control Traffic in the Lightpath Termination Procedure

Similar to the lightpath establishment case, the control traffic incurred in the lightpath termination procedure is investigated by observing the respective operations executed by the controller on ROADMs and transponders. Nevertheless lightpath termination is a simpler and shorter procedure compared to the establishment, especially in ROADMs.

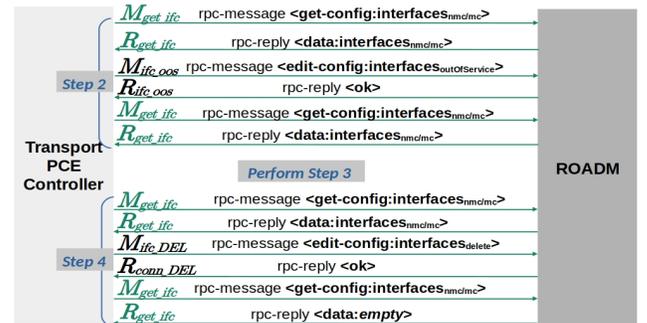
1) *Controller-ROADM Traffic*: During lightpath termination, the controller reverts back the ROADM configurations performed during lightpath establishment. The control operations involved in this termination procedure mainly include - (i) the manipulation and deletion of the lightpath and (ii) the manipulation and deletion of the logical interfaces created within the ROADM. As described in [3], these operations involve the following sequence of actions:

- Step 1: Setting the connection power to the minimum acceptable value of -60 dBm.
- Step 2: Setting the admin state of the interfaces to *out-of-service*.
- Step 3: Deleting the lightpath between these interfaces.
- Step 4: Deleting the logical interfaces.

Lightpath deletion is performed in steps 1 and 3 while steps 2 and 4 are executed to remove the logical interfaces. Note that the controller validates and verifies each operation by requesting the actual state of the corresponding configuration from the ROADM before and after every execution. Figure 5 summarizes the exchange of signalling messages during the



(a) Manipulation and deletion of a lightpath.



(b) Manipulation and deletion of logical interfaces.

Fig. 5. NETCONF operations: controller-ROADM.

execution of these four steps. As seen in Fig. 5a, the controller modifies the power level as per Step 1 and verifies it by requesting the lightpath state. Similarly, after modifying the admin state of the corresponding logical interfaces (as per step 2), the controller requests the lightpath state before performing the delete operation and verifies it after the operation is executed. Likewise, as seen in Fig. 5b the current state of the logical interfaces are fetched each time, before and after, every `<edit-config>` operation - as highlighted in green. From this analysis, it follows that the total size of the controller-to-ROADM messages is given by:

$$\begin{aligned} S(M_{ROADM_d}) &= 4 * S(M_{get\_conn.}) \\ &+ S(M_{conn\_min}) + S(M_{conn\_DEL}) \\ &+ 4 * S(M_{get\_ifc}) + S(M_{ifc\_oos}) \\ &+ S(M_{ifc\_del}) + 12 * S(M_{TEnvv}) \end{aligned} \quad (17)$$

Based on the message sizes in Table I, the total size of the controller-to-ROADM and ROADM-to-controller messages during lightpath termination are  $S(M_{ROADM_d}) = 41.18$  KB and  $S(R_{ROADM_d}) = 31.56$  KB, respectively.

2) *Controller-Transponder Traffic*: The controller creates logical interfaces on both the network-side and the client-side ports of a transponder while setting up a lightpath - recall (14) and Fig. 4. During lightpath termination, the controller restores the transponder configuration as it existed before the lightpath setup. The main sequence of actions performed during this procedure, as described in [3] and observed in our simulation testbed, is:

- Step 1: Setting the transmit power of the OCH interface - created with the request message  $M_{och\_net}$  in (14) - to a minimum value of -5 dBm and setting the admin state to *out-of-service*.
- Step 2: Setting the equipment-state of the circuit pack that supports the OCH interface to *not-reserved-available*.
- Step 3: Setting the admin state of both the network-side and the client-side logical interfaces of the transponder to *out-of-service*. Recall that these interfaces are created with the request messages  $M_{odu\_net}$ ,  $M_{otu\_net}$  and  $M_{eth\_client}$  in (14).
- Step 4: Deleting the *out-of-service* interfaces.
- Step 5: Finally deleting the OCH interface.

Similar to the ROADM operations in Fig. 5, the controller verifies and validates each step by fetching the actual state of each configuration from the transponder before and after the step execution. For example, in step 1, the controller requests for the actual state of the OCH interface using  $M_{get\_och}$  before and after manipulating both the transmit power and the admin state through the message  $M_{och\_min}$ . From the sequence of messages exchanged throughout the execution of the aforementioned five steps the total size  $S(M_{XPDR_d})$  of the controller-to-transponder messages sent during lightpath

termination can be derived as follows:

$$\begin{aligned} S(M_{XPDR_d}) &= 4 * S(M_{get\_och}) + S(M_{och\_min}) \\ &+ 2 * S(M_{get\_cp}) + S(M_{cp\_nra}) \\ &+ 4 * S(M_{get\_ifc}) + S(M_{ifc\_oos}) \\ &+ S(M_{ifc\_del}) + M_{och\_del} \\ &+ 15 * S(M_{TEnvv}) \end{aligned} \quad (18)$$

From Table. I, we get  $S(M_{XPDR_d}) = 30.09$  KB. The corresponding total size of the transponder-to-controller messages sent during lightpath termination is  $S(R_{XPDR_d}) = 27.60$  KB.

#### IV. CONTROL TRAFFIC INDUCED BY DYNAMIC LIGHTPATH SETUPS AND TERMINATIONS

The analysis in Section III shows that the controller follows identical operations with all ROADMs across the lightpath and similarly with all transponders. In fact, the number of NETCONF messages and their sizes are identical for all ROADMs and likewise equivalent for all transponders. Because OpenROADM does not standardize the controller implementation, a differently implemented controller could result in minor variations in the control traffic compared to Transport PCE. This is due to the fact that the sequence of operations could slightly differ from controller to controller. Nevertheless, every controller has to follow the same procedures like Transport PCE to be interoperable with any OpenROADM device.

From the equations derived in Section III, it follows that the establishment and termination of a single lightpath that traverses  $R$  ROADMs and uses  $X$  transponders imposes the following control traffic volume  $S(M_L)$  between the controller and the data plane:

$$\begin{aligned} S(M_L) &= \sum_{r=1}^R \{S(M_{ROADM_c}) + S(M_{ROADM_d})\} \\ &+ \{S(M_{XPDR_c}) + S(M_{XPDR_d})\} * X \end{aligned} \quad (19)$$

Similarly, in the opposite direction, the control traffic volume  $S(R_L)$  between the data plane and the controller is:

$$\begin{aligned} S(R_L) &= \sum_{r=1}^R \{S(R_{ROADM_c}) + S(R_{ROADM_d})\} \\ &+ \{S(R_{XPDR_c}) + S(R_{XPDR_d})\} * X \end{aligned} \quad (20)$$

In a dynamic T-SDN network, optical connection setups and terminations occur continuously and can be modelled as a Poisson process. Assuming a lightpath carried load of  $x$  Erlang on a network element (i.e. ROADM and/or transponder), and a lightpath holding time of  $t_H$  seconds, the average controller-to-ROADM and controller-to-transponder signalling traffic - in Mb/s - that stems from setups and terminations of lightpaths is:  $(x/t_H) * \{S'(M_{ROADM_c}) + S'(M_{ROADM_d})\}$  and  $(x/t_H) * \{S'(M_{XPDR_c}) + S'(M_{XPDR_d})\}$ , respectively. The control traffic in the reverse direction, i.e. from a device to the controller, can be simply obtained by replacing  $R_X$  by  $M_X$ . Note that, as the NETCONF messages are transmitted via TCP [11], a TCP/IP header overhead of at least 40 Bytes must be added to the respective NETCONF message sizes

yielding  $S'(M_{ROADM_c})$ ,  $S'(M_{ROADM_d})$ ,  $S'(M_{XPDR_c})$ ,  $S'(M_{XPDR_d})$  and  $R'(M_{ROADM_c})$ ,  $R'(M_{ROADM_d})$ ,  $R'(M_{XPDR_c})$ ,  $R'(M_{XPDR_d})$  to obtain the correct bitrates.

By using these expressions, we now assess both the control traffic to/from a single ROADM that faces a traffic load of  $x$  Erlang as well as the control traffic to/from the respective transponders. Optical connections are assumed to have a holding time of  $t_H = 600$ s and the carried load  $x$  is varied between 10 and 1000 Erlang. Figure 6 depicts the overall control traffic between the controller and the transponders that results from the establishment and termination of lightpaths dependent of the carried load  $x$ . As observed, the controller-transponder traffic grows linearly as the load  $x$  increases with constant slopes of 0.67 kb/s per Erlang (controller-to-transponder traffic) and 0.78 kb/s per Erlang (transponder-to-controller traffic), respectively. As discussed in Section III, each NETCONF RPC request message is acknowledged by a corresponding reply message from the device, and therefore, the number of requests and replies are identical. Thus, the rate of controller-to-transponder *request* messages per second equals the rate of transponder-to-controller *reply* messages per second. The dependency of these message rates on the carried load  $x$  is shown in Fig.7.

Concerning a ROADM, we evaluate the dependency of the

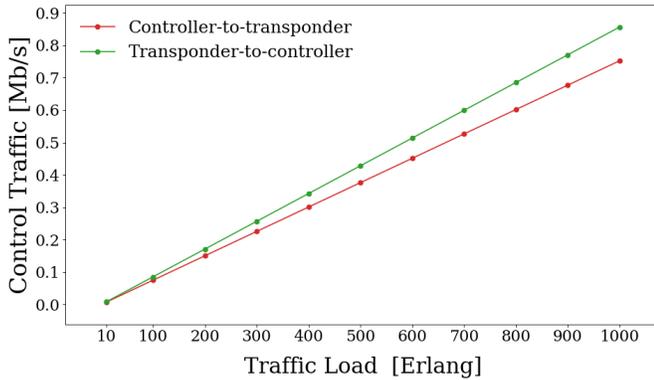


Fig. 6. Controller-to-transponder and transponder-to-controller traffic.

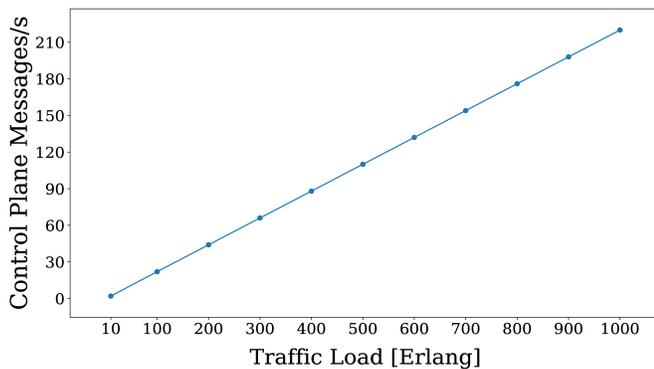
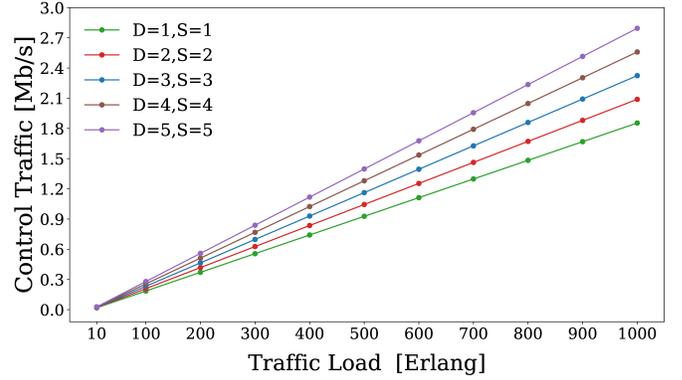
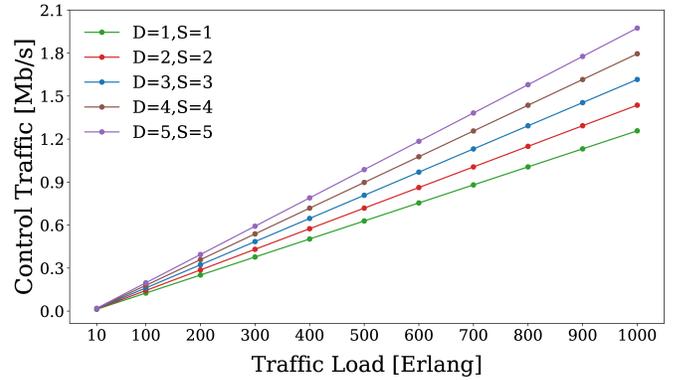


Fig. 7. Number of messages per second between the controller and transponders.

controller-ROADM signalling traffic on the load  $x$ , the degree  $D$  and the number of SRGs  $S$  of the ROADM. The ROADM-to-controller and the controller-to-ROADM traffics are plotted in Fig. 8a and Fig.8b, respectively. For the ROADM,  $D = S$  is assumed, which corresponds to a typical configuration where each degree is installed with an add/drop module that - besides adding and dropping lightpaths - may also support regeneration through back-to-back transponder configurations. It can be seen that the difference between any two curves



(a) ROADM-to-controller traffic.



(b) Controller-to-ROADM traffic.

Fig. 8. Control traffic between the controller and the ROADM.

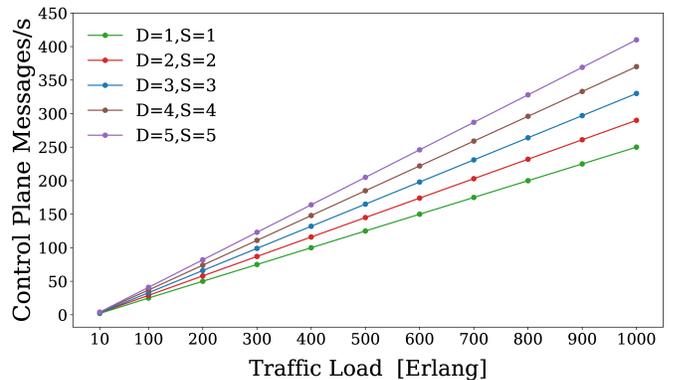


Fig. 9. Number of messages per second between the controller and the ROADM.

with distinct  $D$ ,  $S$  values grows as the load  $x$  increases. For low  $x$  values, the differences are quite marginal. Similar to Fig. 7, Fig. 9 shows the number of messages per second sent from the controller towards the ROADM dependent of the load  $x$  for different values of  $D$  and  $S$ . For example, in the case of 1000 Erlang, and for  $D = S = 5$ , the controller sends approximately 420 *request* messages/s and also has to process the corresponding *reply* messages from the ROADM at the same rate. These message rate results can be utilized to estimate the CPU processing requirements of the controller for handling the signalling traffic.

A general observation valid for both ROADMs and transponders is that the device-to-controller traffic is higher than the controller-to-device traffic. The former either corresponds to responses to controller requests which mainly contain XML encoded device information related to  $\langle$ get-config $\rangle$  operations, or corresponds to simple acknowledgments related to  $\langle$ edit-config $\rangle$  transactions. This implies that the size of response messages that contain device data is larger than the corresponding  $\langle$ get-config $\rangle$  request messages issued by the controller. On the contrary, the size of the acknowledgements is smaller than the corresponding  $\langle$ edit-config $\rangle$  messages issued by the controller to create or update device configurations. However, in Section III it has been discussed that the controller fetches the device configuration not only for the initial device recognition but also for the validation and verification of each write operation, implying that the number of  $\langle$ get-config $\rangle$  requests outnumbers the  $\langle$ edit-config $\rangle$  requests, thereby resulting in a higher device-to-controller traffic.

## V. CONCLUSION

In this paper we analyze and model the control traffic that stems from the setup and termination of lightpaths in Open-ROADM/Transport PCE compliant optical networks. By means of an example we evaluate the dependency of the control traffic between a controller and a ROADM and the respective transponders on the optical connection traffic load. Furthermore the impact of the ROADM internal configuration on the control traffic is investigated. Our results could assist future research on the design of DCNs considering scalability and reliability aspects.

## REFERENCES

- [1] A. Kumari and A. S. Sairam, "A Survey of controller placement problem in software defined networks," arXiv:1905.04649[cs.NI], 2019.
- [2] R. Alvizu, G. Maier, N. Kukreja, A. Pattavina, R. Morro, et al., "Comprehensive survey on T-SDN: software-defined networking for transport networks," IEEE Comm. Surveys & Tutorials, pp. 2232 -2283, June 2017.
- [3] Open ROADM MSA, "Open ROADM MSA release 7 device white paper v1.2.pdf," [Online]. Available: <http://www.openroadm.org/download.html>, December 2020.
- [4] K. Choumas, D. Giatsios, P. Flegkas and T. Korakis, "The SDN control plane challenge for minimum control traffic: distributed or centralized?," 16th IEEE Annual CCNC, February, 2019.
- [5] M. Karatisoglou, K. Choumas and T. Korakis, "Controller placement for minimum control traffic in OpenDaylight clustering," 2019 IEEE 2nd 5G World Forum (5GWF), November 2019.
- [6] A. Bianco, P. Giaccone, A. Mahmood, M. Ullio and V. Vercellone, "Evaluating the SDN control traffic in large ISP networks," IEEE International Conference on Communications (ICC), September 2015.
- [7] R. Casellas, R. Vilalta, R. Martínez and R. Muñoz, "SDN Control of Disaggregated Optical Networks with OpenConfig and OpenROADM," 10th ONDM Conference, pp. 452-464, February 2020.
- [8] A. Triki, C. Betoule, G. Thouenon, O. Renais, N. Pelloquin, G. Lambert, et al., "OpenROADM compliant SDN controller for a full interoperability of the optical transport network," 2018 ECOC, pp. 1-3, November 2018.
- [9] TransportPCE, "Transport PCE documentation," [Online]. Available: <https://docs.opendaylight.org/projects/transportpce/en/latest/index.html>.
- [10] NTS, "Network topology simulator," [Online]. Available: <https://github.com/Melacon/ntsim/tree/master/ntsimulator>, August 2020.
- [11] RFC 7589, "Using the NETCONF Protocol over Transport Layer Security (TLS) with Mutual X.509 Authentication," [Online]. Available: <https://datatracker.ietf.org/doc/html/rfc7589>, June 2015.

# Design and performance characteristics of extended C-band amplifiers

Lutz Rapp

ADVA Optical Networking SE

Märzenquelle 1–3

98617 Meiningen, Germany

lrapp@adva.com

**Abstract**—Enlarging the wavelength range used for data transmission is the most economic step for further increasing the capacity of optical transmission links. Such an extension can either be achieved by using a single amplifier with larger bandwidth or by using two amplifiers in parallel. For a typical two-stage design, the single device solution with increased bandwidth is compared with the split-band approach using two amplifier devices in parallel in various scenarios and in view of different parameters. In particular, the dependence of noise figure on gain bandwidth and measures for improvement are presented. It is furthermore considered up to which capacity the use of a single EDFA is advantageous.

## I. INTRODUCTION

Increasing the capacity of optical networks is an important prerequisite for introducing new applications. Among several options, enlarging the bandwidth available for data transmission is considered as the most economic solution for the near future since it is compatible with the installed fiber infrastructure [1].

Signal amplification in optical networks is mainly achieved by means of erbium-doped fiber amplifiers (EDFAs) [2] that have reached a very mature level of development with low failure rates. Although this technology is able to cover wavelength ranges of 80 nm width and more, the bandwidth of commercial amplifiers is typically limited to around 35 nm [3] for performance reasons.

Separate amplifiers designed for the conventional wavelength band (C-band) and the long wavelength band (L-band), respectively, are used in parallel for increasing capacity [4]. However, such a solution entails additional challenges with regard to the design of optical transmission systems and their control.

Not all links in an optical network need to be upgraded to the full capacity offered by the two wavelength bands in the next time such that extending the gain range of C-band EDFAs to longer wavelengths is sufficient for a longer time period while avoiding many of the drawbacks of a split-band solution (setup [B] in Fig. 1). However, such amplifiers are known for increased noise figures [5].

For a typical two-stage design [6], the single device solution (setup [A]) is compared with the split-band approach using

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016663, B5G-open

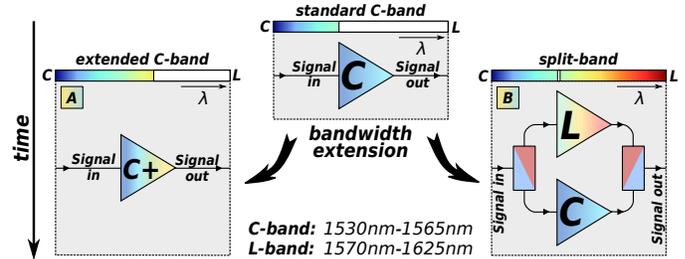


Fig. 1: Capacity extension of existing C-band links by using an extended C-band amplifier (setup [A]) or by parallel amplification of signals in two wavelength bands (setup [B])

two amplifier devices in parallel in various scenarios and in view of different parameters. In particular, the dependence of noise figure (NF) on gain bandwidth and measures for improvement are presented. Furthermore, it is shown that the reported increase of the noise figure can be avoided by appropriate amplifier designs. Based on simulation results, it is furthermore considered up to which capacity the use of a single EDFA is advantageous.

## II. THEORETICAL CONSIDERATIONS

Different gain profiles of an EDFA can be produced by varying the average population probability of the metastable

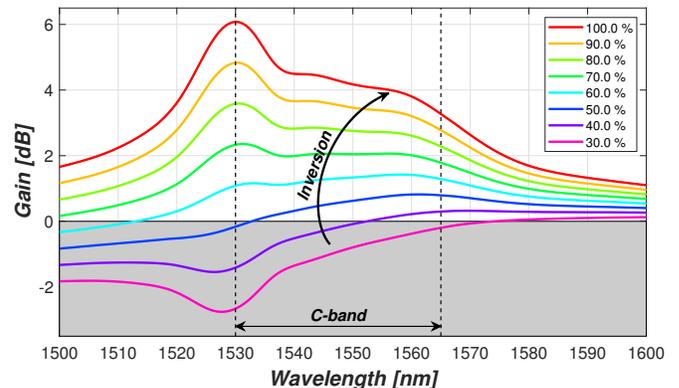


Fig. 2: Gain profile of a commercial EDF for different average values of the inversion per 1 m length

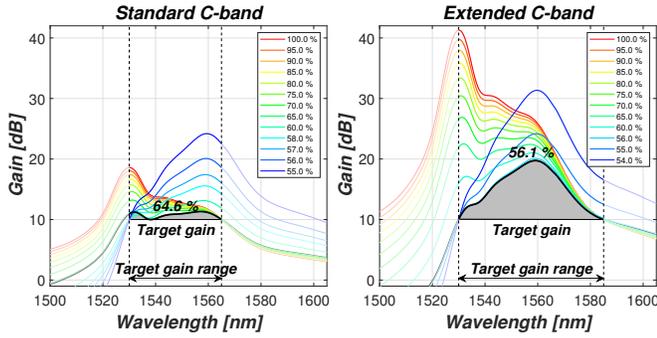


Fig. 3: Determining the total length of the EDF for different gain profiles in order to guarantee a minimum gain at each wavelength within the target gain range. Left side: standard C-band from 1530 nm to 1565 nm. Right side: extended C-band from 1530 nm to 1585 nm

level, usually called inversion. Examples of the resulting gain spectrum are illustrated in Fig. 2 for several inversion levels. When designing a broadband EDFA for wavelength-division multiplexing (WDM) applications, typically a gain profile providing minimum gain variation within the target wavelength range for signal amplification is selected for efficiency reasons, and the amplifier control is designed to maintain this gain profile under all operating conditions. Furthermore, a gain-flattening filter (GFF) [7] compensating for the resulting gain variations within the target wavelength range is embedded in the amplifier setup.

At a given total gain adjusted by the gain control of the amplifier, average inversion and the resulting gain profile are governed by the total length of the EDF coils. In practice, using longer fiber lengths for providing a predefined gain leads to smaller average inversion. In the design phase of an amplifier, the total length of the EDF coils is defined in such a way that the total gain provided by the EDF coils is equal to or larger than the target gain across the entire target wavelength range for the desired average inversion. This requirement comes from the fact that correcting gain differences by a GFF only allows to reduce gain. For a target gain of 10 dB, this aspect is illustrated in Fig. 3 for various average inversions and

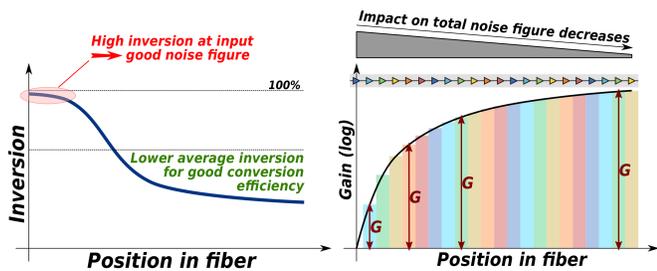


Fig. 4: Left side: typical inversion profile of a codirectionally pumped EDF. Right side: model describing an EDF as a cascade of erbium-doped fiber segments of infinitesimal length

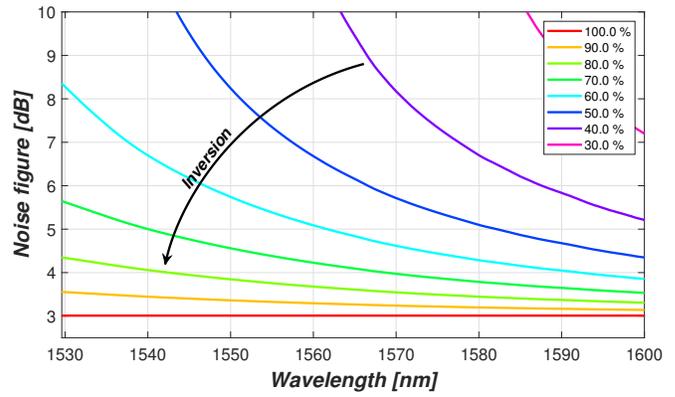


Fig. 5: Noise figure versus wavelength for various values of constant inversion

for different optical bandwidths. The gain profiles providing minimum gain variations are represented by black curves. Obviously, enlarging the gain bandwidth requires to lower the average inversion and to accept larger gain variations to be compensated.

Inversion along the axis of a codirectionally pumped erbium-doped fiber (EDF) decreases with increasing distance from the input, as illustrated on the left side of Fig. 4. The right side of this figure shows a model describing the EDF as a cascade of amplifier segments of infinitesimal length with constant inversion, wherein the cumulative gain increases continuously from one segment to the next segment. Thus, the contribution of these infinitesimal segments to the total noise figure of the amplifier decreases continuously along the cascade of amplifier segments according to Friis's formula [8] [9] and the noise figure is predominantly determined by the first amplifier segments. This allows to gain further insight into the wavelength dependence of the noise figure across the target wavelength range by considering the noise figure of an amplifier with constant inversion as displayed in Fig. 5 for different levels of the inversion. At full inversion, the contribution of such an amplifier segment is independent of wavelength, whereas lower inversion mainly affects shorter wavelength channels. In conclusion, the noise figure of C-band channels is much more sensitive to a reduction of the inversion at the input of an EDF as compared with the higher wavelength channels in the L-band.

### III. BANDWIDTH EXTENSION SCENARIOS

In principle, the target wavelength range of a C-band EDFA can be extended towards shorter and longer wavelengths. However, the potential of an extension towards shorter wavelengths is quite limited due to the steeper decrease of the gain with decreasing wavelength starting at the lower border of the C-band. In contrast, extending the target gain range to larger wavelengths is more promising. In order to cover both options, three different scenarios for extending the gain range are considered that are illustrated in Fig. 6. It is common to all scenarios that the upper limit of the target wavelength range is extended to longer wavelengths in steps of 100 GHz (0.8 nm).

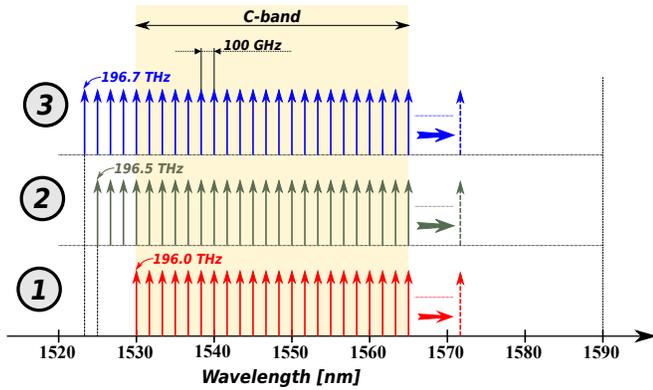


Fig. 6: Different scenarios for increasing the gain range of amplifiers

In scenario ①, the target wavelength range starts at 196.0 THz (1529.4 nm) which is typical for many C-band designs. In contrast, scenarios ② and ③ also support amplification below the typical C-band and the target wavelength ranges start at frequencies of 196.5 THz (1525.7 nm) and 196.7 THz (1524.1 nm), respectively. In the following, simulation results are presented for signals distributed with an equal channel spacing of 100 GHz over the available bandwidth.

Noise figure of a multi-stage amplifier comprising several stages is governed by the noise figure of the individual stages and the signal power distribution across the amplifier setup, in particular at the input of the different stages. Except for the first stage, the input power into the stages strongly depends on the attenuation of the embedded gain-flattening filter (GFF). As a rule of thumb, smaller maximum attenuation of the GFF allows to achieve smaller noise figures. For the different scenarios, Fig. 7 displays on the left side the maximum attenuation of the GFF versus upper limit of the target wavelength range. The right side shows the same attenuation values versus width of the gain range with a strong increase

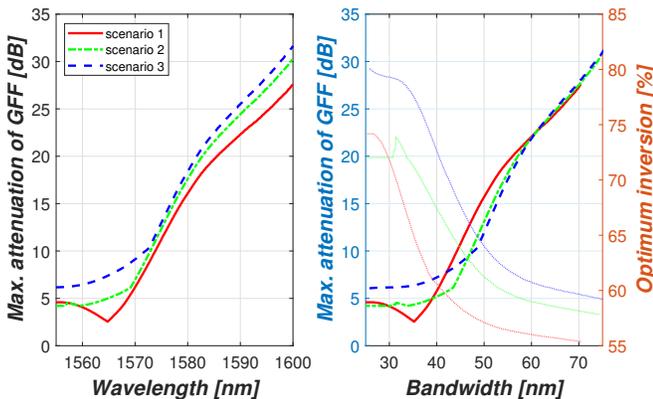


Fig. 7: Maximum attenuation of a single gain-flattening filter used in setup [A](#) versus upper limit of the gain range (left side) and the resulting bandwidth (right side, left axis). Dotted curves indicate optimum inversion versus bandwidth

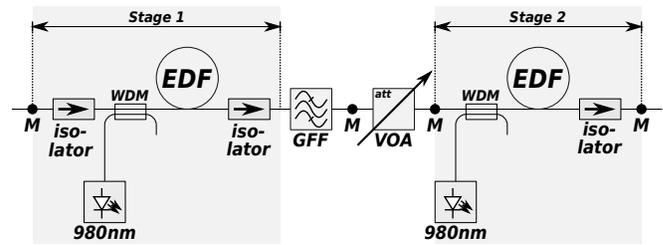


Fig. 8: Typical amplifier setup with one gain-flattening filter (setup [A.1](#))

with increasing bandwidth for all scenarios if the bandwidth exceeds 40 nm. Dashed curves referring to the right axis indicate the optimum inversion level determined as described above and for which the spectral attenuation profile of the GFF has been determined. For all scenarios, the optimum inversion decreases with increasing bandwidth.

#### IV. IMPACT OF WAVELENGTH RANGE ON NOISE FIGURE

The setup of medium performance amplifiers for coherent transmission typically comprises two stages with an intermediate variable optical amplifier (VOA) and one GFF, as illustrated in Fig. 8 (setup [A.1](#)). High output powers are achieved by pumping each of the stages with a separate pump laser diode. Nowadays, fiber coupled pump powers up to 950 mW are available at pump wavelengths around 980 nm [1]. Furthermore, a copropagating pumping scheme used in both stages guarantees optimum noise figure.

With a two-stage setup, the overall noise figure for output powers above 20 dBm strongly depends on the maximum pump power available for the first stage. The main reason is that lower pump powers result in smaller gain of the first stage, whereas the fact that only lower inversion is achieved with smaller pump powers is of minor importance, as confirmed by the noise figure values indicated for the first stage only (see Figs. 9 and 10). Achieving optimum noise performance furthermore requires to adapt the relative length

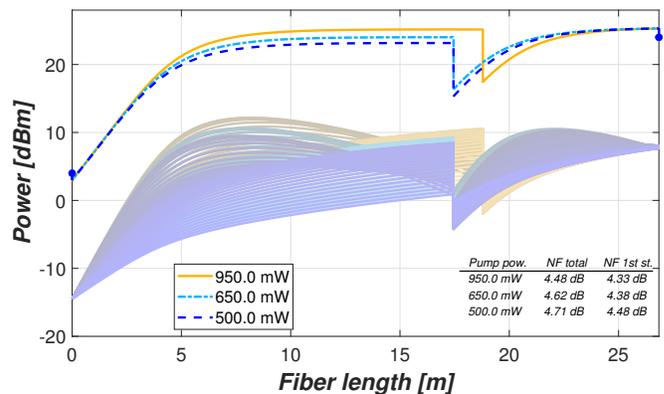


Fig. 9: Power distribution in an erbium-doped fiber amplifier designed for a gain range up to 1575 nm with an output power of 24 dBm and a gain of 20 dB (setup [A.1](#), scenario ①)

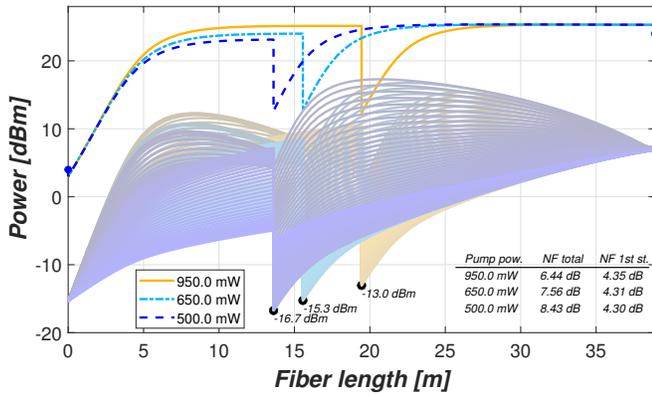


Fig. 10: Power distribution in an erbium-doped fiber amplifier designed for a gain range up to 1585 nm with an output power of 24 dBm and a gain of 20 dB (setup [A.1](#), scenario [1](#))

of the involved EDF coils to the available pump powers. For a maximum wavelength of 1575 nm using scenario [1](#), the power distribution within an amplifier yielding optimum noise performance is illustrated in Fig. 9 for three different pump power levels of the first stage, wherein the upper three curves indicate the total signal power at each location and the lower curves illustrate power evolution for each wavelength channel. Since the input power into the second stage is for all channels and for all pump powers by at least 10 dB higher than at the amplifier input, the overall noise figure only marginally depends on the pump power.

Similar plots for a by 10 nm larger bandwidth, i.e. for a target wavelength range up to 1585 nm, are presented in Fig. 10. Due to the larger maximum attenuation of the GFF, the input powers into the second stage are significantly smaller and similar to the input power. This leads to a strong degradation of the noise figure and its strong dependence on the pump power of the first stage. With 950 mW pump power, the noise figure equals 6.44 dB, but increases to 7.56 dB and 8.43 dB for pump powers of 650 mW and 500 mW, respectively.

Repeating the described design optimization and analysis for the different scenarios with varying bandwidth and for different output powers leads to the results shown in Fig. 11. The plots indicate the maximum noise figure observed within the target wavelength range for a maximum pump power of 950 mW supplied to the first stage. Obviously, extending the target wavelength range to smaller wavelengths (scenarios [2](#) and [3](#)) provides better noise performance as compared with just extending the C-band to larger wavelengths according to scenario [1](#) for bandwidths up to around 58 nm, whereas scenario [1](#) is advantageous for larger bandwidths. Due to the dominating impact of the power distribution within the amplifier setup, stronger increase of the noise figure with increasing bandwidth is observed for larger output powers.

For scenario [1](#), corresponding data have also been determined for a pump power of 500 mW provided to the first stage. For comparison purposes, the results are displayed in Fig. 12

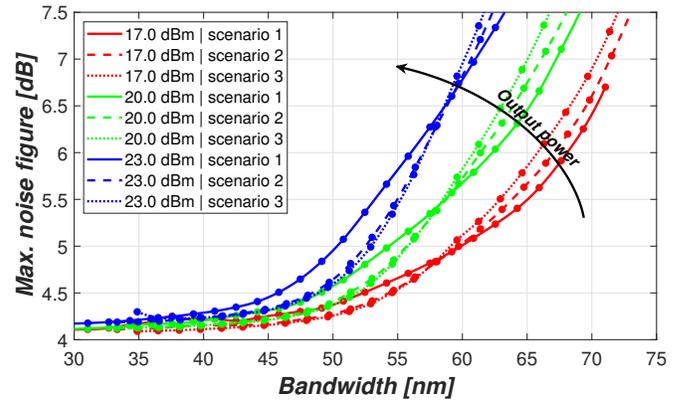


Fig. 11: Noise figure versus bandwidth for different signal output powers and scenarios using setup [A.1](#) (gain: 20 dB)

together with the results for 950 mW pump power launched into the first stage. At equal bandwidth and output power, the difference of the results for the two considered pump powers is quite small for bandwidths below 45 nm but may even exceed 1 dB for bandwidths around 65 nm and high output powers.

Amplifiers for the L-band can be designed to provide noise figures that are comparable with C-band amplifiers at least at larger gain values [10] and for wavelengths up to 1605 nm. At a first glance, this seems to turn into a significant advantage in favor of the split-band setup [B](#) (see Fig. 1). However, the whole setup comprising both amplifiers, the C/L-band splitter and the C/L-band combiner needs to be considered as a single amplifier when comparing the two setups with regard to system performance. In fact, the passive loss introduced in the signal path by the C/L-band splitter placed in front of the amplifiers leads to a noise figure degradation in both bands. Typically, such band-splitters attenuate the signals by around 1 dB. In consequence, the noise figure of the overall setup is by 1 dB larger than the noise figure of the individual amplifiers.

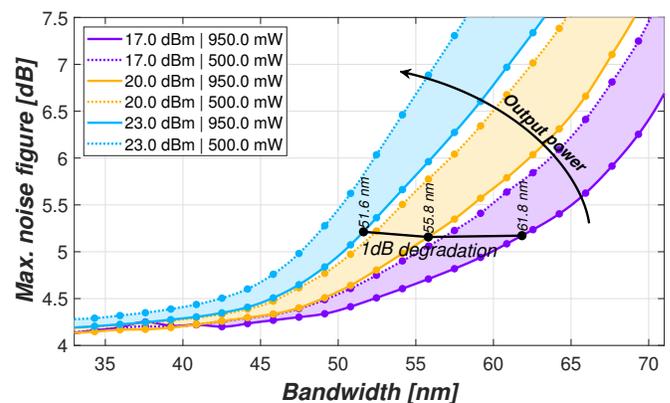


Fig. 12: Noise figure data showing the impact of the maximum pump power of the first stage for various signal output powers for scenario [1](#) and setup [A.1](#) (gain: 20 dB)

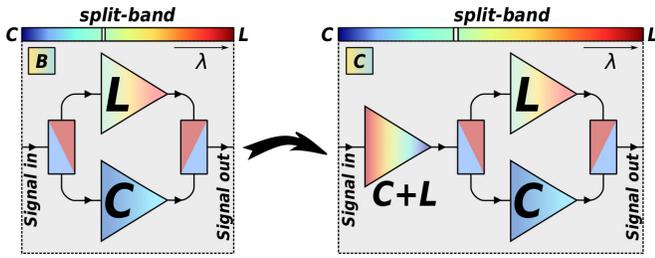


Fig. 13: Extension of the split-band setup [B] (left side) by a common first amplifier stage for better noise figure (setup [C], right side)

The black line in Fig. 12 indicates at which bandwidth the maximum noise figure of an extended C-band amplifier has increased by 1 dB as compared with a reference amplifier designed for the C-band only. In conclusion, the extended C-band design provides better system performance than the split-band setup for bandwidths up to 62 nm at an output power of 17 dBm and for bandwidths up to 52 nm at an output power of 23 dBm. On the other hand, the results also show that extended C-band designs are not suitable for providing larger bandwidths up to 90 nm that are achievable with split-band designs.

This advantage of the extended C-band amplifier would be nullified if the impact of the losses of the C/L-band splitter on noise figure of the split-band setup is avoided. Placing a first amplification stage amplifying both bands in front of the C/L-band splitter [11] in setup [B] is a promising approach to achieve such an improvement, which leads to setup [C] shown on the right side of Fig. 13. For the following analysis, data transmission in the wavelength range from 1529.6 nm to 1564.7 nm (C-band) and in a wavelength range of equal width from 1572.1 nm to 1609.2 nm (L-band) is investigated.

A major design challenge of setup [C] becomes clear from the power distributions for the different wavelength channels illustrated in Fig. 14, wherein C-band channels are represented

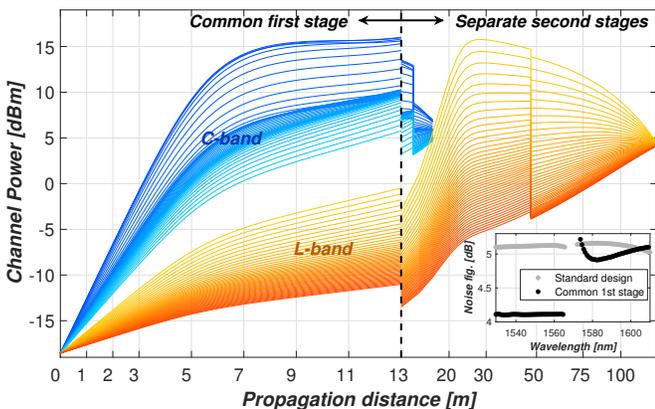


Fig. 14: Power distribution of C-band and L-band channels in an amplifier design with a common first stage for both bands (setup [C]). Deformed x-axis for propagation distances longer than 13 m

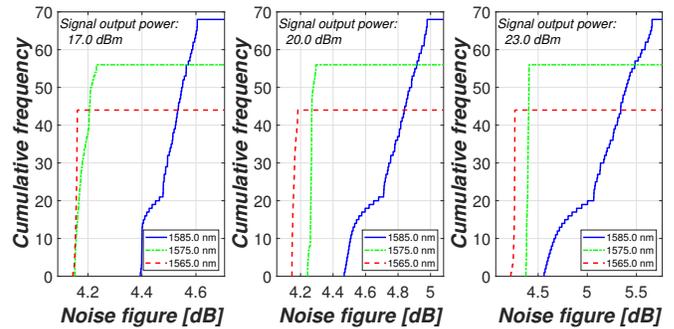


Fig. 15: Cumulative distribution of the noise figure for various signal output powers. Parameter: upper limit of the gain range using scenario ①

by blue colors and red colors refer to L-band channels. For better readability, the x-axis has been deformed for propagation distances after the C/L-band splitter, i.e. where the different wavelength bands are amplified in separate EDF coils of different length. Since the C-band channels experience significantly larger gain than the L-band channels in the first stage, the gain of the L-band channels produced before the C/L-band splitter is not sufficient to render the contribution of this passive component negligible according to Friis's formula.

Black dots in the inset of Fig. 14 indicate noise figure across both bands for setup [C] with a common first stage, whereas gray dots represent the noise performance for setup [B] with two completely separate amplification branches. Obviously, using a common first stage leads to improved noise figure in the C-band, but there is no benefit in the L-band. In summary, both setups are very similar with respect to the maximum noise figure value arising within the target wavelength range of operation. It remains that extended C-band amplifiers are the better choice for required bandwidths up to around 60 nm at smaller output powers and up to 50 nm at higher output powers as compared with split-band approaches designed for providing amplification over 70 nm or more.

## V. NOISE FIGURE DISTRIBUTION

Noise performance of optical amplifiers is typically characterized by indicating the maximum noise figure value within the target wavelength range. However, this parameter only provides information about the channel that is most affected by amplified spontaneous emission (ASE) generated in the amplifier under consideration. For the present work, it is of particular importance to know whether the increased maximum noise figure pertains to the added wavelength range only or if the noise level is also increased in the wavelength range that has already been supported before extension. For this purpose, the distribution of the noise figure values for signal channels with an equal channel spacing of 100 GHz is considered in the following. Please note that the total number of channels depends on the target bandwidth and is therefore different for the three different configurations considered in the following.

The cumulative frequency of occurrence is a suitable parameter for characterizing the distribution of the noise figure

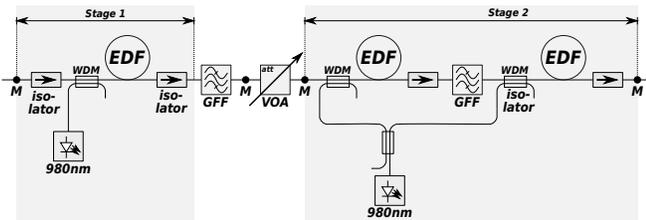


Fig. 16: Two-stage amplifier setup with two gain-flattening filters (setup [A.2])

among the channels and is shown in Fig. 15 for three different upper limits of the target wavelength range according to scenario ①. The maximum pump power provided to the first stage amounts to 950 mW and the amplifier is operated at a gain of 20 dB with output powers of 17 dBm, 20 dBm, and 23 dBm. Following the curves from smaller noise figures to larger noise figures, the curve for larger bandwidth is always found below the curve for the smaller bandwidth until the latter reaches its maximum. It follows that extending the target wavelength range towards larger wavelengths does not only affect the added channels but also increases the noise level for the wavelength channels that have been transmitted before the bandwidth extension.

## VI. SETUP FOR IMPROVED NOISE FIGURE

Results presented so far clearly indicate that the increase of the maximum noise figure of extended C-band amplifiers is predominantly caused by the power distribution over the amplifier setup and not by the physical properties of the energy level transitions the amplification mechanism is based on. Thus, tailoring the power profile across the amplifier setup is a promising approach to avoid deterioration of the noise performance with increasing bandwidth. Such an optimization of the power profile can be realized by spreading the attenuation required for gain flattening over several GFFs.

An amplifier setup derived from setup [A.1] but comprising two GFFs and analyzed in the following is illustrated in Fig. 16 (setup [A.2]). In order to accommodate the second GFF, the EDF coil of the second EDFA stage is split into two parts. Furthermore, both parts are supplied with pump power emitted by a single pump laser diode by means of pump power splitting. For simplicity reasons, both GFFs use the same filter shape providing half of the attenuation of the GFF in setup [A.1].

Distribution of the total signal power (upper curves) as well as the power of the individual wavelength channels (set of lower curves) is displayed in Fig. 17 for 950 mW pump power launched into the first stage and for an upper limit of the target wavelength range of 1585 nm. For comparison purposes, power profiles for setup [A.1] already known from Fig. 10 are also shown in this figure. With the described modification, the minimum signal power level after the first stage is by almost 10 dB higher as compared with setup [A.1] so that the noise figure of the overall amplifier is only 0.28 dB higher than the noise figure of the first stage for an extended wavelength range

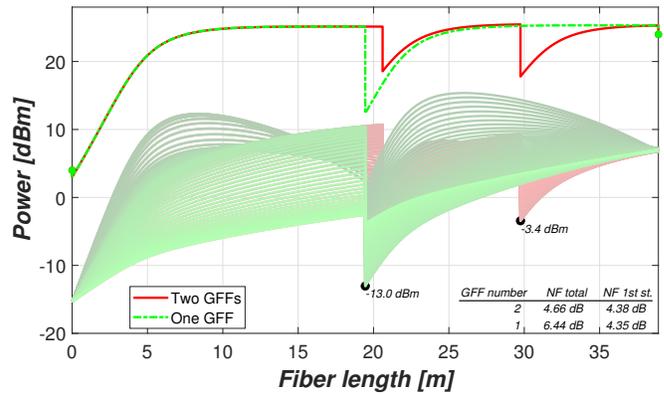


Fig. 17: Power profiles in EDFA setups using one (setup [A.1]) or two (setup [A.2]) gain-flattening filters, respectively, for scenario ① with a maximum wavelength of 1585 nm. Pump power in first stage: 950 mW, Gain: 20 dB

up to 1585 nm. Furthermore, the noise figure of this optimized setup [A.2] is by 1.78 dB better as compared with the default setup [A.1] comprising two stages and one GFF.

Of major interest is the comparison of this design optimized for an extended C-band up to 1585 nm with an EDFA design according to setup [A.1] that is conceived for the standard C-band. For a fair comparison, this EDFA is operated at the same per channel power and thus reduced total output power. Specifically, the total output power is set to 22 dBm in contrast to the 24 dBm of the modified setup for the extended C-band which yields a maximum noise figure of 4.18 dB in this operating point. In consequence, extending the available bandwidth by 20 nm leads to an increase of the maximum noise figure by 0.48 dB. Optimizing the filter function of the GFFs is expected to further decrease the noise figure. Thus, the modified setup allows for an increase of the bandwidth by 60% with only minor degradation of the noise figure. However, a disadvantage of the approach is the increased cost caused by additional components such as an additional filter and an additional WDM coupler, and by the higher pump power needed for pumping the second stage.

## VII. CONCLUSIONS

Presented results reveal that the noise figure degradation associated with erbium-doped fiber amplifiers (EDFAs) operating in an extended C-band are almost completely caused by the power profile in standard EDFA designs. The extent of the degradation does not strongly depend on the lower limit of the target wavelength range for signal transmission. However, degradation of the noise figure can be strictly limited by using a modified EDFA design with more than one gain-flattening filter. Simulation results indicate that increasing the bandwidth by 20 nm as compared with the standard C-band is possible with a quite small increase of the maximum noise figure by 0.5 dB. However, the higher capacity comes at the expense of higher cost.

## REFERENCES

- [1] L. Rapp and M. Eiselt, "Optical amplifiers for multi-band optical transmission systems," *Journ. Lightwave Technol.*, vol. 40, pp. 1579–1589, 2021.
- [2] P. C. Becker, N. A. Olsson, and J. R. Simpson, *Erbium-Doped Fiber Amplifiers — Fundamentals and Technology*. San Diego, London, Boston, New York, Sydney, Tokyo, Toronto: Academic Press, 1999.
- [3] G. P. Agrawal, *Fiber-Optic Communication Systems*. New York: John Wiley & Sons, 1997, ch. 8.
- [4] S. Yamamoto, H. Taniguchi, Y. Kisaka, S. Camatel, Y. Ma, D. Ogawa, K. Hadama, M. Fukutoku, T. Goh, and K. Suzuki, "First demonstration of a C+L band CDC-ROADM with a simple node configuration using multiband switching devices," *Optics Express*, vol. 29, pp. 36 353–36 365, 2021.
- [5] ZTE Corporation, "The information about the extended C-band," International Telecommunication Union, Geneva, Switzerland, ITU-T T17-SG15 Contribution 2059, Sep. 2020.
- [6] S. Kinoshita, T. Okiyama, and K. Moriya, "Wideband WDM erbium-doped optical fiber amplifiers for 10 Gb/s, 32-channel SMF transmission system," *Fujitsu Sci. Tech. J.*, vol. 35, pp. 82–90, 1999.
- [7] D. E. M. Albaa, J. M. C. Herrera, and G. A. P. Leguizamón, "Design of a flattening filter using fiber Bragg gratings for EDFA gain equalization: An artificial neural network application," *Ciencia e Ing. Neogran.*, vol. 29, pp. 25–36, 2019.
- [8] L. Rapp, "Investigation into the significance of failure rate calculations for erbium-doped fiber amplifiers in view of the reliability of communication networks," *Eur. Trans. Telecommun.*, vol. 20, pp. 580–593, 2008.
- [9] K. Grobe and M. Eiselt, *Wavelength Division Multiplexing — A Practical Engineering Guide*. Hoboken, New Jersey: John Wiley & Sons, 2014.
- [10] L. Rapp, "Noise figure limit of L-band EDFAs," *AEÜ Int. J. Electron. Commun.*, vol. 56, no. 2, pp. 109–114, 2002.
- [11] Y. Sun, J. W. Sulhoff, A. K. Srivastava, J. L. Zyskind, C. Wolf, T. A. Strasser, J. R. Pedrazzani, J. B. Judkins, R. P. Espindola, A. M. Vengsarkar, and J. Zhou, "An 80 nm ultra-wide band EDFA with low noise figure and high output power," in *Europ. Conf. Opt. Comm.* Edinburgh, UK: IEE, 1997, pp. 69–72.

# Energy Efficiency Limits to ICT Equipment Lifetime

Klaus Grobe  
Global Sustainability  
ADVA Optical Networking SE  
Martinsried, Germany  
[KGrobe@ADVA.com](mailto:KGrobe@ADVA.com)

**Abstract**—Circular economy (CE) has severe implications on electronics systems design. Maximum re-use has to be considered in the initial product design phase. For certain ICT products, this may be contradicted by specific technical and operational aspects. The environmental impact of certain classes of ICT equipment is clearly dominated by use-phase energy consumption, which can be derived from lifecycle analyses (LCA). This necessitates a focus on energy efficiency, which may be conflicting with design in support of CE. Increasing energy efficiency of successor product generations can set an upper limit to maximum lifetime. Above this limit, further use of old equipment becomes net-negative, e.g., for the global warming potential, according to LCA. A use-phase / production-phase emissions ratio is defined that allows identifying whether replacement of ICT or other electronic equipment for reasons of energy efficiency and reduced total-lifetime emissions is necessary. From that, near-optimum lifetime and near-optimum replacement cycles, if existent, can be derived. Several examples are provided. High-capacity, fiber-optic WDM transport equipment should be replaced, due to fast pace of energy-efficiency gain. Other ICT equipment like fiber-plant monitoring devices should not be replaced for this reason.

**Index Terms**—Lifetime optimization, lifecycle assessment.

## 1. INTRODUCTION

THE EU Horizon-2020 project C-SERVEES [1] investigates improvements of circular economy business models (CE BMs) in the electrical and electronics sector. This includes theoretical analyses of possible CE BMs and four practical demonstrators with different equipment classes that aim at proving at least some of the respective CE BM findings. This has been complemented by extensive analyses regarding lifetime optimization of electronic equipment, aiming at total-lifetime emissions minimization. These analyses are based on lifecycle assessments (LCA), together with predictions of electricity emission factors and the development of energy-efficiency gain over the next three decades. It also considers major product characteristics like the average use mode.

CE has severe implications on product design and lifetime requirements. For certain ICT products, these may be contradicted by certain technical and operational aspects. This is particularly true for ICT infrastructure (i.e., core-network)

products. These products typically show very strong use-phase LCA dominance, which has an impact on their lifetime or the CE aspect of longevity.

In this paper, an optimum-lifetime analysis will be given. This analysis covers a generalization to other classes of electronic equipment.

## 2. ICT ENVIRONMENTAL STATUS

The main components of the ICT sector, or the Internet, are wired (access and backbone) and wireless (access only) networks, data centers, and end-user equipment. The networks split into backbone or core and access (incl. wireless) parts. The core networks consist of aggregation switches, routers and fiber-optic WDM long-distance transport. For these infrastructure equipment classes, some 80-90% of the environmental impact are determined by the use-phase energy consumption, which can be derived from lifecycle analyses [2]-[4]. This is particularly true for the global warming potential (GWP), that is, emissions of greenhouse gases (GHG, i.e., carbon dioxide (CO<sub>2</sub>), methane (CH<sub>4</sub>), nitrous oxide (N<sub>2</sub>O), and ozone (O<sub>3</sub>)).

Part of the use-phase dominance is owed to the typical use mode of ICT infrastructure equipment, which is 24/7 always-on, at high utilization, and over long lifetime (around 8 years or more). The second relevant reason is the ICT bandwidth or bitrate growth, which is forecasted to persist also for the next couple of years [5]-[7]. This is shown in Fig. 1, together with the resulting global ICT energy consumption according to [8], [9] and the associated GWP.

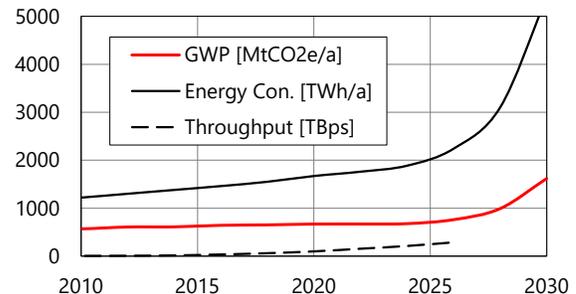


Fig. 1. ICT throughput based on [5]-[7] (extrapolated to 2026), and related growth in energy consumption [8], [9] and resulting emissions (GWP)

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 776714.

The red curve in Fig. 1 shows the resulting GWP. Beyond 2020, these emissions may lead to total global GWP increase despite the necessity to massively decrease emissions, following the Paris Agreement [11]. This is driven by bandwidth-hungry applications like video downloads, 5G, 6G, big data, and others.

For calculating the GWP, average electricity emission factors that are linearly decreasing from 0.4 kgCO<sub>2</sub>e/kWh in 2020 to 0.3 kgCO<sub>2</sub>e/kWh in 2030 have been used.

So far, the main ecodesign effort for ICT infrastructure equipment went into the increase of power efficiency. As an example, the power efficiency for WDM transponders improved from ~10 W/Gbps in the beginning in the mid-90s to around 0.2 W/Gbps for the latest generation. At the same time, the throughput (e.g., total WDM system transmission capacity) did cope with global ICT throughput, as shown in Fig. 2.

Here, WDM system capacity and ICT global throughput are displayed in ordinate log scale. Both curves have similar slope. Similar slope can also be derived for other infrastructure equipment (switches, routers) [13]. This means that the amount of equipment and associated raw-material consumption and waste electrical and electronic equipment (WEEE) is growing slower than its energy consumption.

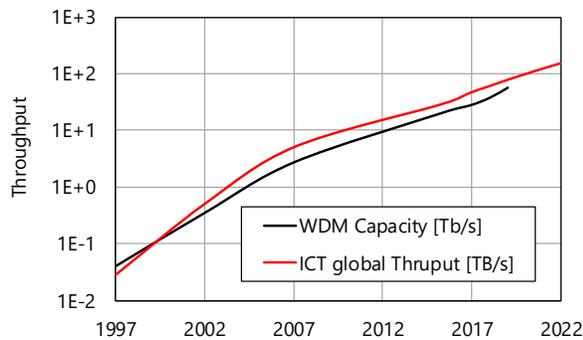


Fig. 2. Power-efficiency development for WDM [15] (top) and Internet throughput increase [5]-[7] and WDM system-capacity increase [15] (bottom)

The development of the EU WEEE generation is shown in Fig. 3 [12]. It can be seen that WEEE generation has a moderate, almost linear, growth rate.

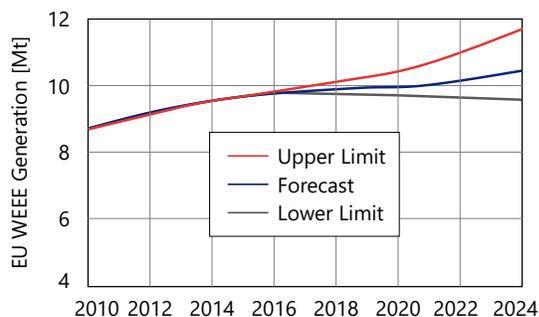


Fig. 3. Development of WEEE generation in the EU [12]

WEEE is one of the main aspects to be improved by CE.

CE also aims at minimizing raw-material intake. This is relevant for the ICT sector when considering materials regarded scarce or critical. From the EU list of critical raw materials 2017, at least one half of the elements are critical for the ICT sector [14].

As a first conclusion, components and raw-material aspects are important for ICT infrastructure gear and equipment with similar use mode, but less important than energy-efficiency. Whether this persists will be analyzed in the next chapter.

### 3. RELEVANT PARAMETERS

#### 3.1 Consideration of the GWP

When calculating optimum lifetime of ICT and other electronics equipment, suitable measures must be used. For our calculations, we use the LCA midpoint parameter of the global warming potential, GWP. Due to global warming and the urgently required actions to reduce carbon emissions, GWP is possibly the most important LCA midpoint impact category today. However, circular economy and its related business models are primarily considering general resource depletion. This covers fossil resources and thus, indirectly, emissions. In addition, and more importantly on the long run, it also covers other abiotic resources, including critical raw material.

For the ICT products that are considered in this chapter primarily, GWP is nonetheless a valid parameter for lifetime optimization. This follows from product LCA that considers various midpoint parameters.

Two LCA examples are shown in Fig. 4 for the TeraFlex system. TeraFlex is an example of latest-generation coherent WDM transport equipment [3]. Since the typical use mode is 24/7 always-on, the system is clearly dominated in GWP (and other mid-point LCA parameters) by its use phase. Note that in this case, absolute power consumption is high, but at the same time, TeraFlex is one of the most power-efficient WDM systems globally. Here, power efficiency is rated in W/Gbps. ICT equipment like routers and core switches has similar characteristics, albeit at somewhat higher power consumption per throughput.

We show two LCAs, one calculated using average EU electricity grid mix, the second one using 100% renewable energy. This is relevant since the related emission factors (with dimension [kgCO<sub>2</sub>e/kWh]) obviously have high influence on LCA. Considering both emission factors (EU grid mix and fully renewable) is also relevant to compile meaningful results for the complete period up to 2050, when Carbon neutrality is targeted in the EU and when electricity must be based on renewable energy.

In the upper part of Fig. 4, the TeraFlex LCA is shown for grid-mix electricity. All midpoint categories shown are dominated by the use phase. In particular, abiotic resource depletion has a result similar to GWP. The same can be shown for Ethernet switches and IP routers [4]. This necessitates continued work toward better energy efficiency since use-phase impact is driven by the related energy consumption. In

turn, this may be conflicting with design in support of CE, at least in those cases where total development effort is limited.

When renewable energy is used, things change but use phase is still mostly dominant, see the lower part of Fig. 4. The reason is that we use a non-zero emission factor of  $EF_{RE} = 0.04 \text{ kgCO}_2\text{e/kWh}$  for renewable energy (RE). This averaged value considers complete lifecycle electricity emissions. These cover a large range due to differences in technology, local resource conditions and methodological approaches for assessment [16]. Wind, solar or water energy is not fully emissions-neutral since the construction of the respective plants causes emissions.

The share of renewable technologies and their emission factors of the future energy market are not yet known. For example, lifecycle GHG emissions are in the range of 7-56  $\text{gCO}_2\text{e/kWh}$  for wind energy and 5-217  $\text{gCO}_2\text{e/kWh}$  for photovoltaic. Producing renewable energy is never carbon-neutral unless compensation techniques or techniques like CCS (carbon capture and storage) are used [17].

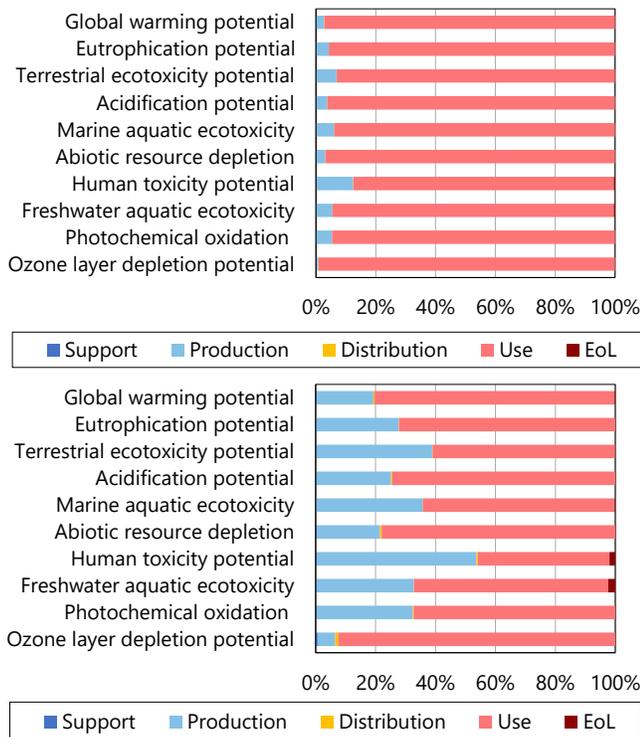


Fig. 4. LCA for the TeraFlex system using grid-mix electricity (top) and renewable energy (bottom).

Note that also for using renewable energy, abiotic resource depletion performs similar to GWP. Therefore, GWP is considered a valid measure for lifetime optimization and is used hereinafter.

LCA has been conducted in line with ISO 14040/44 and ETSI EN 203 199, with openLCA, together with theecoinvent and GaBi electronics data bases, emission factors and use mode as discussed and a lifetime assumption of 8 years.

### 3.2 Parameter extrapolation

In order to derive optimum lifetime of an electronic device, several steps are required. First, we perform the derivation based on total resulting GWP, that is, emissions. This way, the production phase with its impact on raw material can be traded off against the use phase, which can lead to high emissions over lifetime. We do not consider the other lifecycle phases here since typically, they are far less relevant in LCA. This simplification is certainly true for infrastructure ICT equipment, refer to Fig. 4.

Next, average emission factors need to be extrapolated into the future since we need to consider potentially long or extended lifetime that can go beyond 2040 or so. We use two sets of emission factors (EF) that are extrapolated up to 2050. The first set assumes a linear decline in the – average – EF toward 100% renewables in 2050.

The second EF set is more aggressive than the first one. It reaches 100% RE in 2044. Although it might become necessary in the near future, even faster approach toward 100% RE is not seen realistic today on a global scale. Note that the EF are average, that is, global figures.

Also note that the extrapolated EF are in line with the B2DS scenario of the IEA [18] and the RCP2.6 scenario of the IPCC [19], that is, scenarios that support less than 2°C global warming compared to the pre-industrial state.

The resulting emission factors are shown in Fig. 5.

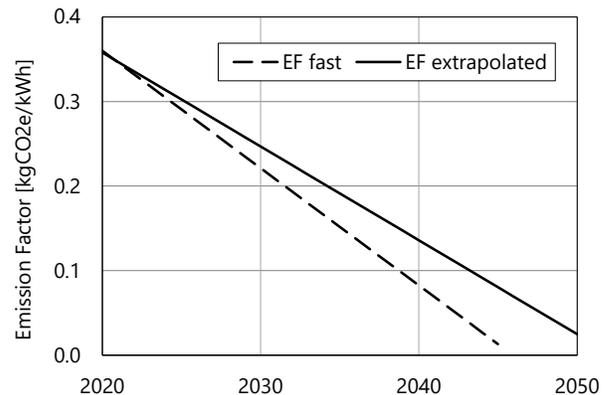


Fig. 5. Extrapolation of emission factors. The dashed curve comprises faster declining values.

The next parameter that is relevant for calculating GWP into the future is the development of the gain in energy efficiency in electronics and photonics. It is obviously relevant for ICT gear and other electronics equipment. The problem is that this gain started saturating since total efficiency is approaching some fundamental physical limits [13]. At today's pace, any gain in electronic-components efficiency must come to a stop in around two decades, given that no disruptive technologies will be found. Also in [13], it is stated that in the last two decades or so, the yearly gain in energy efficiency for main classes of ICT infrastructure equipment (switches, routers, WDM transmission gear) was in the range of 14%. Therefore, we assume non-linear decrease of the yearly efficiency gain from 14% in 2020 toward almost

zero in 2050. This may be regarded pessimistic, but it is not. For example, it is clear that somewhere below 5 nm CMOS (complementary metal-oxide semiconductor) feature size, the CMOS development will have to come to an end.

The resulting development in efficiency is shown in Fig. 6.

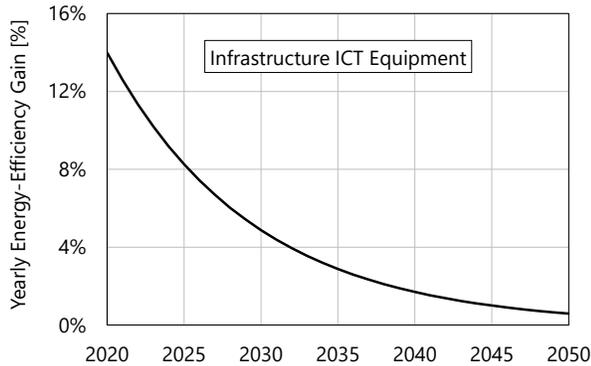


Fig. 6. Extrapolation of yearly gain in electronics energy efficiency.

Fig. 6 can be used to calculate the improvement in efficiency that a successor generation of a piece of equipment can offer after, say, 10 years. Obviously, this improvement depends on absolute time (it is decreasing over time).

#### 4. LIFETIME OPTIMIZATION

##### 4.1 ICT core-network systems

Based on the emission factors and energy-efficiency development discussed in the previous chapter, we calculated total-lifetime GWP for several ICT devices in several scenarios, with and without replacement for reasons of energy efficiency.

As the first device, we analyze TeraFlex, whose LCA was already shown in Fig. 4. We compare a scenario where the system is installed and activated in remains in service for very long time, irrespective of potentially changing service requirements. Then, different scenarios are investigated where the original unit is replaced by successor units that are more efficient according to Fig. 6. For all scenarios, the most relevant emission contributions – use phase and production – are considered, other lifecycle phases (transportation, end of life and contributions from our sites) only have a very small impact and can be neglected.

First, we assume grid-mix electricity and its development of the emission factors as per Fig. 5, extrapolated case. The result for TeraFlex is shown in the upper part of Fig. 7. Three scenarios are shown: no-replacement, replacement every 8 years and a replacement scheme that is close to optimum. This shows that regular replacement, e.g., every 8 years, becomes less efficient over time. Faster replacement at the beginning might be better, but toward 2050, it performs worse. For TeraFlex, close-to-optimum replacement is achieved after 4 years, another 4 years and 8 years, respectively. Replacement every 12 years or slower is not fast enough, and replacement every 4 years also perform worse than the optimum replacement.

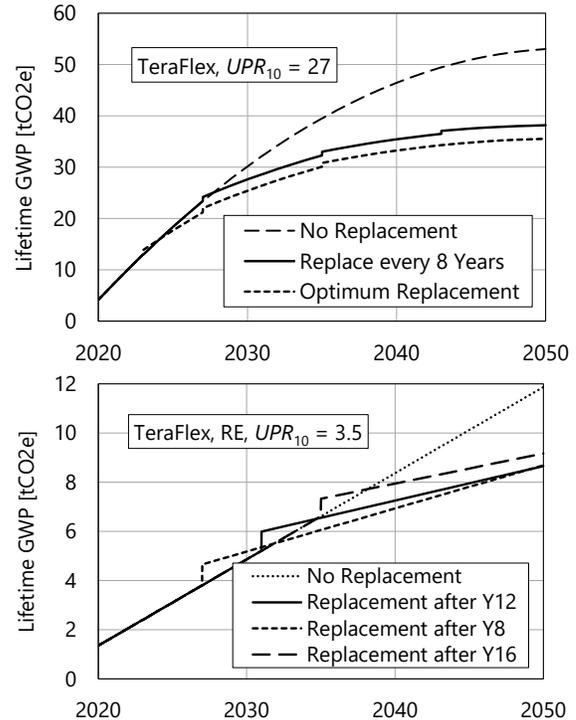


Fig. 7. Replacement scenarios for TeraFlex for grid-mix electricity (top). Replacement for TeraFlex, when using 100% renewable energy (bottom). The steps in the curves represent successor-unit production emissions.

In the top figure, the curves are flattening due to the improving emission factors. The saturation of efficiency gain, together with emission factors that are improving over time, lead to the fact that replacement becomes less frequently necessary over time. In general, availability of successor products as frequent as every 4 years is unrealistic for infrastructure ICT equipment. In addition, efficiency gain becomes smaller with faster replacement. For these reasons, we investigated combinations of 4, 8, 12, and 16 years as replacement periods.

Next, we varied the emission factors to see how stable results are. For this, we used the faster development of the emission factors from Fig. 5. The result does not differ significantly from the previous one.

We also approximated optimum TeraFlex replacement when using 100% renewable energy, see the bottom part of Fig. 7. Due to the low emission factor, only one replacement should take place in the period up to 2050. Replacements after 8 and 12 years respectively achieve the same minimum resulting GWP. Replacement after shorter or longer periods lead to higher total GWP.

Also note that considering renewable energy and related emissions yields the correct results for ICT network operators that already run their entire network with 100% renewable energy. This holds for large network operators like British Telecom or Deutsche Telekom.

We kept the production emissions constant, over time and with changing emission factors. This is not fully exact, but

prediction how production LCA changes with changing emission factors is difficult. Production GWP certainly will get better over time, but the extent is unclear. We regard this a secondary effect which does not have significant influence on our results. Note that keeping the original high production GWP does penalize the more frequent replacement of products. For TeraFlex, the crossover point in time would only come even earlier.

The next photonic ICT device, ALM-64, is an ICT infrastructure, fiber-plant monitoring system. It permanently monitors fibers for their integrity, therefore the use mode is 24/7 always-on again. Technically, the ALM is an optical time-domain reflectometer. Moreover, the ALM is designed for lifetime in excess of 10 years, therefore, its energy consumption is relevant. The LCA of the ALM again shows use-phase dominance. For grid-mix electricity, the use phase caters for 55-90% of the different LCA midpoint parameters.

The lifetime GWP of two ALM scenarios is shown in Fig. 8. In the first scenario, the ALM is installed in the respective infrastructure and remains there. In the second scenario, the original ALM is replaced, after 10 years, by a successor product. Again, we assume that the successor has better energy efficiency after 10 years according to Fig. 5. This translates to a drop in energy consumption – for the same application! – by 55%. This is somewhat optimistic since, e.g., a successor might have been developed five years after the original product, in which case the efficiency gain is smaller. However, we use this value, an improvement of 55% in energy consumption, to clearly identify lifetime effects.

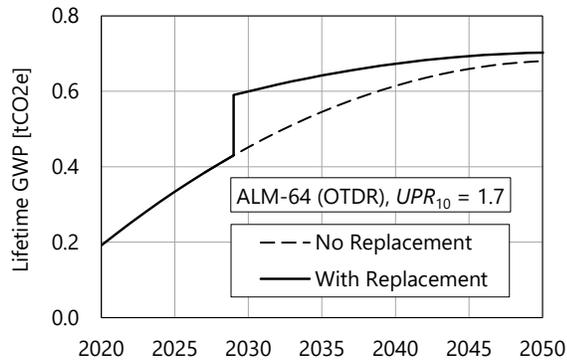


Fig. 8. Lifetime GWP for two ALM-64 scenarios. The steps in the curves represent the (successor) unit production emissions.

In Fig. 8, we consider grid-mix electricity and use-phase and production emissions only. The other lifecycle emissions contributions are negligible, as discussed earlier.

Both curves are flattening over time due to improving emission factors. The curve for the second scenario is even flatter due to improved energy efficiency of the successor. However, the two curves never cross. This means that in most realistic scenarios, the ALM should not be replaced due to improvements in energy efficiency – the ALM consumes sufficiently little energy! In turn this means that the ALM should be (and is) designed for longevity, potential parts reuse and finally recycling.

#### 4.2 Generalization to other (ICT) systems

So far, we investigated two ICT infrastructure products with different characteristics of use versus production impact.

In order to fully generalize our analysis and to more clearly answer the question if and when products should be replaced by more efficient successors, we define the following Use-phase/ Production-phase Ratio  $UPR_{10}$  for the GWP:

$$UPR_{10} := \frac{\text{GWP of the first 10 years in the use phase}}{\text{GWP of the production phase}} \quad \text{Eqn. (1)}$$

$UPR_{10}$  is the ratio of 10 years use-phase GWP and the production GWP according to LCA. Note that *the first 10 years* in use shall be considered since over time, the use-phase GWP decreases thanks to improving emission factors. This allows more meaningful statements for the next two decades. We consider 10 years of use to allow certain averaging of the emission factors and sufficient time for successor product generations to be developed.

Note that  $UPR_{10}$  does depend on the applicable emission factors. Due to the fact that emission factors for grid-mix electricity (Fig. 5) and renewable energy develop differently,  $UPR_{10}$  cannot be normalized to the emission factor but must be considered for different emission factors separately.

In Tab. 1, relevant parameters for the ALM and TeraFlex are summarized.

Tab. 1. GWP parameters of the ALM-64 and TeraFlex

	ALM-64	TeraFlex
Power Consumption	10W	1000 W
Production GWP	160 kgCO <sub>2</sub> e	1.0 tCO <sub>2</sub> e
$UPR_{10}$ Grid Mix	1.7	27
$UPR_{10}$ RE	0.04	3.5

$UPR_{10}$  for the ALM and TeraFlex differs by more than one order of magnitude. Also note the  $UPR_{10}$  differences when using grid-mix emission factors and renewable energy (RE), respectively.

Varying  $UPR_{10}$  across a maximum range and identifying if and when there is a GWP crossover point in time between the two scenarios (no replacement vs. replacement) now allows full generalization beyond ICT products. In the upper part of Fig. 9, this crossover is shown as a function of  $UPR_{10}$  for the case of extrapolated emission factors as per Fig. 5.

The crossover time shows hyperbolic character. Below  $UPR_{10} \approx 2$ , there is no crossover anymore, it moves toward infinite duration. This is obviously the case for ALM, see Fig. 8. This means the ALM is indeed at the one end of the  $UPR_{10}$  scale. For  $UPR_{10} > 20$ , the crossover is within the first year. Significantly higher  $UPR_{10}$  values are unlikely since this would require products with very high power consumption but low production GWP. In this part of the curve, it is clear that products must be replaced if more efficient successors are available. It is also clear that TeraFlex (with  $UPR_{10} = 27$ ) is at the other end of the scale, compared to the ALM.

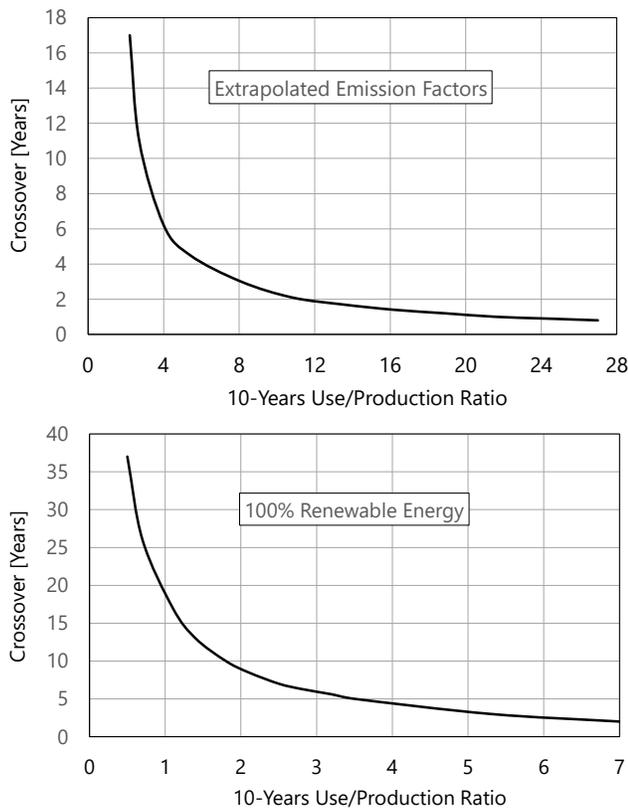


Fig. 9. Crossover time toward total net-positive emissions for emission factors extrapolated up to 2050 as a function of the  $UPR_{10}$  (top). Crossover time toward total net-positive emissions for  $EF_{RE} = 0.04 \text{ kgCO}_2\text{e/kWh}$  for 100% renewable energy as a function of the  $UPR_{10}$  (bottom).

Next, we repeat the crossover calculation for 100% renewable energy. The result is shown in the lower part of Fig. 9. Not surprisingly, the crossover time shifted upward, and with 100% RE, no  $UPR_{10}$  values above 7 were reached. However, even when using 100% RE, the principle characteristic does not change. For products that are massively energy consuming in the use phase, replacement after a certain point in time still makes sense in terms of total resulting GWP. This holds, as discussed, for small non-zero emission factors for RE.

Next, we derived the crossover  $UPR_{10}$  range that separates the replacement from the no-replacement domains. Again, this cannot be derived precisely in general (dependence on  $UPR_{10}$ , dependence on absolute time, dependence on emission factors, unknown development cycles of successor products). However, we can estimate the crossover range for  $UPR_{10}$  that separates the replacement domains. In case a given product falls into this crossover range, more precise calculations must be done.

The estimation can be done by fixing the production GWP or the use-phase power consumption and varying the other of the two parameters to the value where the two scenarios (replacement, no replacement) yield the same GWP. The crossover  $UPR_{10}$  then simply is the ratio of the two resulting parameters as per Eqn. (1).

Then, for the extrapolated emission factors of Fig. 5, a value of  $UPR_{10 \text{ Crossover}} \approx 4$  results. This is in line with the upper part of Fig. 9. There,  $UPR_{10} \approx 4$  is the range where the crossover-time curve most massively changes in slope.

Note that the period beyond 2050 is less relevant because RE emission factors must be assumed. Here, only products with high  $UPR_{10}$  require replacement.

Next, the  $UPR_{10 \text{ Crossover}}$  calculation is repeated for the case of running the equipment with 100% renewable energy. Now,  $UPR_{10 \text{ Crossover}} \approx 0.9$ . Again, this is in line with the bottom part of Fig. 9, where 0.9 is the range where the gradient of the curve clearly changes. Obviously, the smaller the emission factors are, the more the crossover point moves toward higher power consumption relatively to the production impact.

Note that these calculations must be adapted to the emission factors that the respective equipment will see in its future use phase.

In general, products with  $UPR_{10}$  clearly below 4 (grid mix) or clearly below 0.9 (RE), respectively, do not require replacement. Products with  $UPR_{10}$  clearly above 4 (grid mix) or clearly above 0.9 (RE), respectively, do require it. Products with  $UPR_{10}$  near the crossover values 4 (grid mix) or 0.9 (RE), respectively, should be analyzed in more detail to identify the optimum replacement scenario. Similar considerations hold for the preference on energy efficiency in ecodesign. Products with high  $UPR_{10}$  require consideration of energy efficiency first, followed by circular-economy consideration. For products with small  $UPR_{10}$ , it is vice versa.

Further note that  $UPR_{10}$  may have an influence on the choice of the most-suitable product-service system (PSS) [20] as the supporting CE BM. In cases of very high  $UPR_{10}$ , it may be beneficial to select a PSS that supports (improvements of) operational efficiency, namely energy efficiency. In turn, certain PSS can have a massive influence on  $UPR_{10}$ . In particular, PSS that aim at massively improving product utilization, e.g., by better sharing, may also lead to massively increased lifetime energy consumption, thus increasing  $UPR_{10}$ . In these cases, it should be checked whether the respective products need to be replaced because of energy-consumption reasons after certain periods.

In order to verify the generalization via  $UPR_{10}$ , we repeated the analysis for two third-party devices from the data-center space.

The first device is a storage platform for up to 84 drives. Its production GWP is  $5.0 \text{ tCO}_2\text{e}$ , target lifetime is 10 years, and average power consumption is  $1690 \text{ W}$  [21]. From this,  $UPR_{10} = 9.12$  for grid-mix electricity results. This clearly indicates that the storage platform should be replaced somewhere after the target lifetime for reasons of energy efficiency improvements. This is confirmed in Fig. 10.

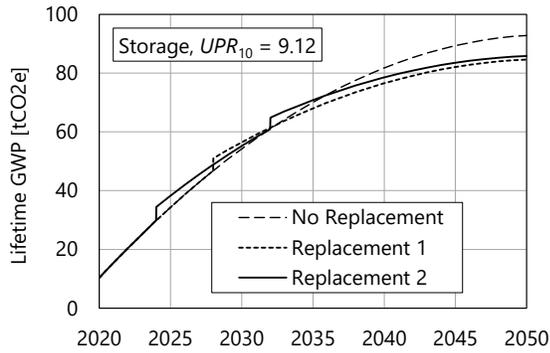


Fig. 10. Lifetime GWP for a storage platform [21]. Two replacement scenarios indicate that the device should be replaced when more efficient successor units become available.

We used somewhat slower efficiency gain here. This considered the effect that efficiency gain of hard disk drives is assumed to be slower than the one of semiconductors as shown in Fig. 6. Nonetheless, a clear indication of replacement for reasons of energy efficiency and related emissions can be derived.

The second device is a high-performance server with maximum configuration of two 28-core CPUs and three GPUs [22]. Target lifetime of this device is four years, production GWP is 4.3 tCO<sub>2e</sub>, and power consumption is 308 W, respectively. This translates to  $UPR_{10} = 1.94$  for grid-mix electricity. The lower use-phase impact results from the average utilization (100% load: 10% of the time, 50% load: 35%, 10% load: 30%, idle mode: 25%). The small  $UPR_{10}$  indicates that the device should not be replaced and could hence be given longer lifetime. This is confirmed in Fig. 11.

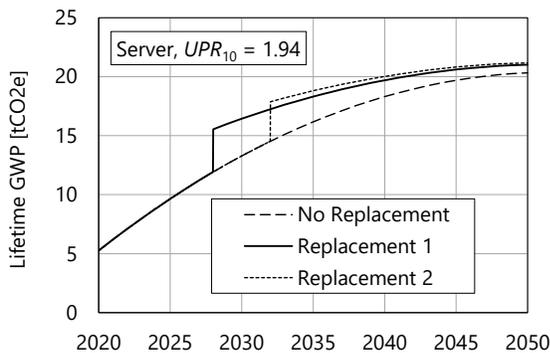


Fig. 11. Lifetime GWP for a server.

Again, we used somewhat slower efficiency gain here, compared to Fig. 6. This considers that the server can be equipped with hard disk drives and that CPU efficiency scaling has already slowed down significantly [23], [24].

#### DISCUSSION AND CONCLUSION

Circular economy aims at reducing raw-material consumption and (electronics) waste generation through the concept of longevity. However, lifetime may be limited, for certain products, by use-phase energy consumption and the efficiency gain that is to be expected by successor products. This effect

is time-dependent since both, the efficiency gain and the electricity emission factors that are to be applied decrease over time.

For a product classification, the use-phase-over-production-phase GWP ratio  $UPR_{10}$  has been defined. When considering average grid-mix emission factors extrapolated into the future,  $UPR_{10} > 4$  indicates the domain where products should be replaced after a certain time when significantly more efficient successor products are available. This lowers total lifetime GWP, including the production GWP of the successor. For  $UPR_{10} < 4$ , the products shall be extended in lifetime, within other limitations like functional obsolescence etc. Even when using renewable energy, this only changes quantitatively. Therefore, replacement requirements will be seen even in two or three decades from now for certain ICT equipment.

Compared to full LCA,  $UPR_{10}$  is simpler (although it requires the production GWP of the device under consideration), and it solves the ambiguity regarding the replacement question that LCA use-phase dominance leaves.

Lifetime GWP optimization using  $UPR_{10}$  may contradict generic CE requirement for longevity. However, in particular as long as average emission factors are clearly above the ones for renewable energy, the product discrimination enabled by  $UPR_{10}$  should be applied in order to avoid net-negative CE rebound effects on global warming. This is particularly true for potential new regulations. These must apply similar electronics products discrimination. Products with very strong use-phase dominance, like TeraFlex, must be allowed to be taken out of service for reasons of energy-related emissions, even if they could be re-used for extended periods from the viewpoint of functionality.

Of course this means that at their end of life, these products have to be recycled to the best-possible extent using best-available technology, in order to recover scarce and precious material. In turn, devices that need not be replaced shall be designed for longevity, including design for maintenance, modularity, high meantime between failures, and finally again design for recycling.

The  $UPR_{10}$  concept can be improved by normalizing  $UPR_{10}$  to the predicted energy-efficiency gain in that period. This yields clearer results in cases where the ICT efficiency gain used in this paper cannot be applied. Similar to the emission factors, the efficiency gain then has to be adopted individually.

#### ACKNOWLEDGEMENT



This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 776714.

## REFERENCES

- [1] [c-serveesproject.eu/](http://c-serveesproject.eu/)
- [2] K. Grobe, "Improved Sustainability in WDM Transport-Network Elements," EGG2016+, Berlin, September 2016.
- [3] K. Grobe and S. Jansen, "Challenges of ICT equipment regarding Circular Economy," EGG 2020+, Session C.2, online, September 2020
- [4] Cisco Corporate Social Responsibility Reports 2019, online: [www.cisco.com/c/dam/m/en\\_us/about/csr/csr-report/2019/pdf/csr-report-2019.pdf](http://www.cisco.com/c/dam/m/en_us/about/csr/csr-report/2019/pdf/csr-report-2019.pdf).
- [5] Cisco White Paper, "The Zettabyte Era: Trends and Analysis," June 2016.
- [6] Cisco White Paper, "The Zettabyte Era: Trends and Analysis," June 2017.
- [7] T. Barnett Jr. et al., "Cisco Visual Networking Index (VNI) Complete Forecast Update, 2017–2022," Cisco Knowledge Network Presentation, Dec. 2018.
- [8] A.S.G. Andrae, "Projecting the chiaroscuro of the electricity use of communication and computing from 2018 to 2030," Preprint Feb. 2019, DOI: 10.13140/RG.2.2.25103.02724.
- [9] A.S.G. Andrae, "Prediction Studies of Electricity Use of Global Computing in 2030," Int. J. Science and Eng. Investigations, Vol. 8, Issue 86, March 2019, ISSN: 2251-8843, Paper ID: 88619-04, pp. 27-33.
- [10] H. Ritchie and M. Roser, "CO<sub>2</sub> and Greenhouse Gas Emissions," 2020, published online at OurWorldInData.org. Retrieved from: [ourworldindata.org/co2-and-other-greenhouse-gas-emissions](http://ourworldindata.org/co2-and-other-greenhouse-gas-emissions).
- [11] United Nations, The Paris Agreement, online: [unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement](http://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement)
- [12] F. Magalini et al. Study on Collection Rates of Waste Electrical and Electronic Equipment (WEEE), possible measures to be initiated by the Commission as required by Article 7(4), 7(5), 7(6) and 7(7) of Directive 2012/19/EU on Waste Electrical and Electronic Equipment (WEEE). Technical Report. 2016. Available online: [ec.europa.eu/environment/waste/weee/pdf/Final\\_Report\\_Art7\\_publication.pdf](http://ec.europa.eu/environment/waste/weee/pdf/Final_Report_Art7_publication.pdf).
- [13] K. Grobe and S. Jansen, "Limits to exponential Internet Growth," EGG2020+, Session A.2, online, September 2020
- [14] European Commission, Study on the review of the list of Critical Raw Materials Executive summary, June 2017. Available online: [ec.europa.eu/docsroom/documents/25421](http://ec.europa.eu/docsroom/documents/25421)
- [15] ADVA WDM systems specifications (partially available under [www.adva.com/en/products](http://www.adva.com/en/products)) and own research. [www.adva.com/en/products/open-optical-transport](http://www.adva.com/en/products/open-optical-transport)
- [16] IPCC, "Energy Systems. In: Climate Change 2014: Mitigation of Climate Change," 2014.
- [17] M. Bui et al., "Carbon capture and storage (CCS): the way forward," Energy & Environmental Science. 11 (5): 1062-1176, 2018. doi:10.1039/C7EE02342A
- [18] OECD/IEA, "Energy Technology Perspectives 2017," ISSN: 2079-2603, ISBN: 978-92-64-27597-3, 2017
- [19] IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp.
- [20] L. Kjær et al., "Product/Service-Systems for a Circular Economy: The Route to Decoupling Economic Growth from Resource Consumption?" J. Indust. Ecology, Vol. 23, No. 1, 2018, pp. 22-25. DOI: 10.1111/jieec.12747.
- [21] OneStor SP2584 Extensible Storage Platform (ESP) Product Life Cycle Assessment (LCA), Seagate data sheet. Available online: <https://www.seagate.com/files/www-content/global-citizenship/en-us/docs/lca-summary-onestor-sp-2584s-20150309.pdf>
- [22] Life Cycle Assessment of Dell PowerEdge R740, Report produced June, 2019. Available online: [https://corporate.delltechnologies.com/content/dam/digitalasset/active/en/unauth/data-sheets/products/servers/lca\\_poweredge\\_r740.pdf](https://corporate.delltechnologies.com/content/dam/digitalasset/active/en/unauth/data-sheets/products/servers/lca_poweredge_r740.pdf)
- [23] J.L. Hennessy and D.A. Patterson, "A New Golden Age for Computer Architecture," Contributed Article, Communications of the ACM, Vol. 62 No. 2, Feb. 2019, pp.48-60
- [24] K. Rupp, "42 Years of Microprocessor Trend Data," Feb. 2018, available online: <https://www.karlrupp.net/2018/02/42-years-of-microprocessor-trend-data/>.

# Two-Sided Endless Optical Polarization Control with 2 x 100 krad/s Tracking Speed

Reinhold Noé<sup>(1,2)</sup>, Benjamin Koch<sup>(1,2)</sup>

(1) Paderborn University, EIM-E, Warburger Str. 100, D-33098 Paderborn, Germany, e-mail: noe@upb.de  
(2) Novoptel GmbH, Helmerner Weg 2, D-33100 Paderborn, Germany, E-mail: info@novoptel.com

**Abstract**—Endless optical polarization control is demonstrated. Polarization transformations before and behind broadband LiNbO<sub>3</sub> polarization transformers fluctuate at about 100 krad/s each, >100 times faster than the slower of these two speeds in earlier two-sided polarization controllers. Mean relative intensity loss during a 115 h test of 2 x 100 krad/s tracking was about 0.007, maximum loss was about 0.04.

**Keywords**— polarization, Lithium Niobate, polarization control

## I. INTRODUCTION

Polarization control has been investigated for coherent optical communication and sensing. It should be endless, i.e. tracking should be possible for polarizations wandering many times around the Poincaré sphere. Rotating fiber loops [1–3], fiber squeezers [4–6] and LiNbO<sub>3</sub> polarization transformers [7, 8] have been proposed decades ago.

The incoming polarization state of a polarization controller or the polarization state analyzed behind it may vary and needs to be tracked. Often, this holds for only one of them. Such one-sided endless polarization control has been reported up to 100 krad/s on the Poincaré sphere with low tracking errors [9].

Until today, those controllers which simultaneously tolerate endless polarization transformation fluctuations before and behind the polarization transformer seem to be quite limited in speed. 200 rad/s tracking behind, concurrently with 15 krad/s before the polarization transformer was reported [10]. To our knowledge, there are no reports of polarization controllers which track fluctuations at both sides where the slower (hence limiting) of these two fluctuation speeds (or any of the two speeds if they are equal) gets near (or above) 500 rad/s. We report two such experiments, using one or two polarization transformers which are LiNbO<sub>3</sub> components.

## II. SETUP: 2 X 100 KRAD/S POLARIZATION CONTROL WITH TWO LiNbO<sub>3</sub> COMPONENTS

Fig. 1 shows the experimental setup. A 1550 nm laser signal is passed to polarization scrambler 1. It consists of 7 rotating electrooptic waveplates: A halfwave plate (HWP) rotating at 100 krad/s, with 3 quarterwave plates (QWP) before and 3 QWPs behind it, all rotating at slow incommensurate speeds. The polarization-scrambled signal is passed through LiNbO<sub>3</sub> polarization transformer 1. At its output there is a polarization beamsplitter (PBS1) which captures the vertically polarized residual power. This is detected by an avalanche photodiode (APD1) and minimized by the controller 1 which sets the LiNbO<sub>3</sub> device voltages appropriately. Horizontal polarization with (almost) full intensity leaves PBS1. This is input to a LiNbO<sub>3</sub> polarization transformer 2. Between its output and PBS2 there is polarization scrambler 2. It is like

polarization scrambler 1 but operates with somewhat different, incommensurate rotation speeds to avoid any control benefits. Controller 2 minimizes the vertically polarized residual power in PBS2, which is detected by APD2. Horizontal polarization with almost full intensity exits from PBS2. Gradient search is applied.

PBS1 converts residual polarization fluctuations into small intensity losses. It thereby decouples very effectively the polarization control processes in controllers 1 and 2. These can run at their target tracking speeds of 100 krad/s each. No time-interleaving or different electrode voltage dither frequencies are needed for controllers 1 and 2.

The multiplication factors of photodetectors APD1, APD2 are set automatically within 7 dB dynamic ranges. There is a small optical attenuation (not depicted) before APD1, in order to let the dynamic ranges of controllers 1 and 2 overlap completely. Apart from this, polarization controllers 1, 2 are similar to that used in [10].

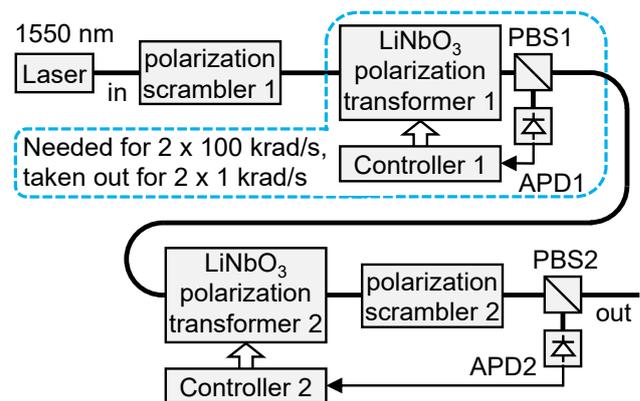


Fig. 1: Setup for control of fast polarization transformation fluctuations at input and at output of polarization controller.

For 2 x 100 krad/s tracking, 2 controllers are needed.

For 2 x 700 rad/s or single-sided 1 x 1000 rad/s tracking, 1 controller suffices.

## III. RESULTS: 2 X 100 KRAD/S POLARIZATION CONTROL WITH TWO LiNbO<sub>3</sub> COMPONENTS

Both polarization controllers accumulate the residual intensity samples into histograms during control. The samples are normalized and referred to as relative intensity error (RIE). Fig. 2 shows the complementary cumulative distribution function  $1-F(\text{RIE})$  of the RIE for both controllers at different nominal scrambling speeds between 200 rad/s and 100 krad/s. Function  $1-F(\text{RIE})$  is the (small) probability of surpassing a certain RIE. At 200 rad/s, maximum observed RIEs were 0.0027 and 0.0031 for polarization controllers 1 and 2 within 5 minutes. At 100 krad/s, maximum RIEs were 0.02 and 0.022 within 5 minutes.

Fig. 3 shows these RIEs, now 0.025 and 0.03, within an extended measurement over 115 hours. Individual RIE results remained unchanged when the other polarization scrambler and the corresponding controller were disabled. This shows that the two control loops are very well decoupled.

The RIEs are calculated from the residual-polarized signal. This means, when looking at the full output signal, the two momentary small RIEs must roughly be added. If we add the individual maximum RIEs 0.025 and 0.03 at  $2 \times 100$  krad/s we reach a sum RIE 0.055. But the two control loops are independent. Therefore total  $1-F(RIE)$  is not so much worse than an individual one because large RIE in one coincides most likely with small RIE in the other polarization control loop. Approximate total RIE is hence calculated by convolving the underlying probability density functions (pdfs), assuming that the two individual RIEs are statistically independent. Result is also plotted in Fig. 3 (red trace). RIE = 0.04 is reached with probability  $<10^{-12}$ .

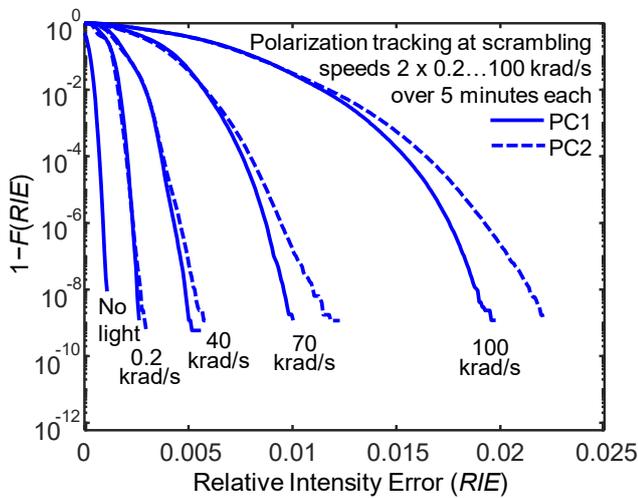


Fig. 2: Complementary cumulative distribution function  $1-F(RIE)$  of relative intensity error ( $RIE$ ), for polarization controllers 1 (PC1) and 2 (PC2) at various speeds.

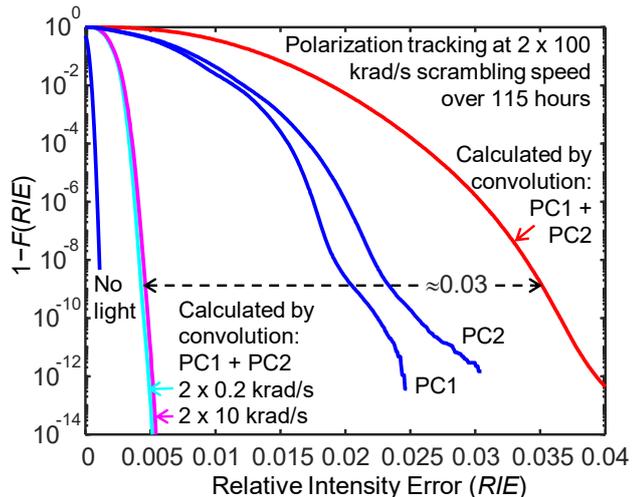


Fig. 3: Complementary cumulative distribution function  $1-F(RIE)$  of relative intensity error ( $RIE$ ), for polarization controllers 1 (PC1) and 2 (PC2) at 100 krad/s (blue). Convolution of underlying probability density function yields calculated  $1-F(RIE)$  of PC1 + PC2 (red/magenta/cyan =  $2 \times 100/10/0.2$  krad/s)

Also shown is  $1-F(RIE)$  calculated by convolutions of the pdfs measured at  $2 \times 10$  krad/s (magenta trace). This speed is not plotted in Fig. 2 because the corresponding traces would partly overlap with those for  $2 \times 0.2$  krad/s.

Fig. 3 likewise shows  $1-F(RIE)$  calculated for  $2 \times 0.2$  krad/s (cyan trace). Horizontal difference between the  $2 \times 0.2/10$  krad/s curves (cyan/magenta) and the  $2 \times 100$  krad/s curve (red) is about 0.03 at  $1-F(RIE) \approx 10^{-9}$ , the lowest relevant for the 5 min measurements.

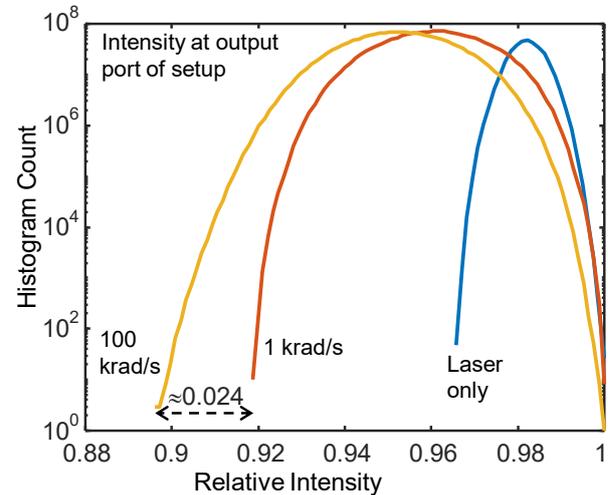


Fig. 4: Intensity histograms at output port of setup (PBS2 out).

In Fig. 4, histogram counts are shown, aligned at a maximum  $RIE = 1$ . The rightmost curve shows a intensity histogram of the laser alone. Intensity drops to about 0.966 in the worst case. Measurement is shorter (1 min) and area under the curve is hence smaller than in the two other curves (5 min each). The middle curve shows a histogram of the intensity at the output of PBS2 during control and 1 krad/s scrambling speed. Lowest intensity is about 0.92, i.e. 0.046 lower than the samples of the laser in a shorter time. This is more than would be expected from the sum of worst RIE at 1 krad/s. Reason is that it also contains the PDL of both scramblers and controllers. The histogram taken at  $2 \times 100$  krad/s shows intensity samples down to 0.896. This drop of another 0.024 is caused by rising RIE at high scrambling speeds. It reasonably agrees with the difference 0.03 of calculated curves in Fig. 3.

Input power of  $\text{LiNbO}_3$  polarization transformer 1 was  $-5$  dBm, permissible range is  $-4 \dots -11$  dBm, or lower if electrical control gain is increased. Attenuations of  $\text{LiNbO}_3$  polarization transformers 1, 2 are 2.2 dB each, plus 0.3 dB for PBS1, together 4.7 dB.

#### IV. SETUP: $2 \times 700$ RAD/S AND $1 \times 1000$ RAD/S POLARIZATION CONTROL WITH ONE $\text{LiNbO}_3$ COMPONENT

Now the experimental setup of Fig. 1 is simplified.  $\text{LiNbO}_3$  polarization transformer 1, PBS1, APD1 and controller 1 are removed and replaced by a through connection. This has the advantage of reduced insertion loss and of course reduced cost. Control algorithm needed to be changed which drastically reduced tracking speed.

## V. RESULTS: $2 \times 700$ RAD/S AND $1 \times 1000$ RAD/S POLARIZATION CONTROL WITH ONE $\text{LiNbO}_3$ COMPONENT

Fig. 5 shows the RIE distribution during tracking with the simplified setup. Maximum RIEs were below 0.015 up to a scrambling speed of about  $2 \times 400$  rad/s. Further increase of scrambling speeds leads to larger maximum RIE, for example 0.03 at  $2 \times 700$  rad/s. Above about  $2 \times 800$  rad/s, the controller occasionally loses track and RIEs become unbounded.

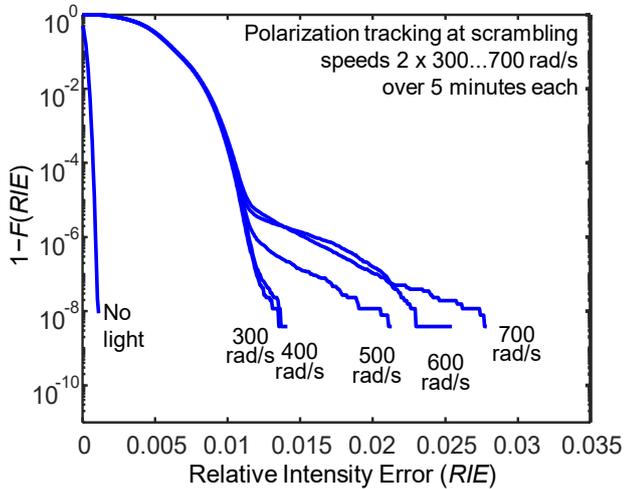


Fig. 5: Two-sided polarization tracking with single  $\text{LiNbO}_3$  component.

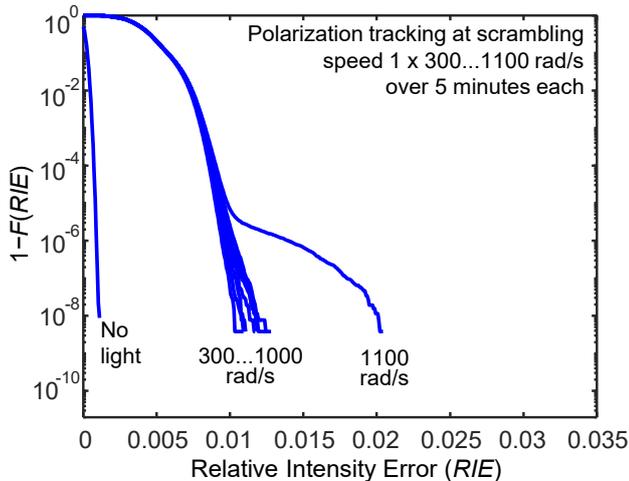


Fig. 6: One-sided (right) polarization tracking with single  $\text{LiNbO}_3$  component.

One-sided polarization tracking was also tested. Fig. 6 right shows good tracking up to  $1 \times 1000$  rad/s, but significantly decreased performance when tracking 1100 rad/s. The changed control algorithm is the reason why this setup is not capable of  $1 \times 100$  krad/s operation like in Section III. and [9].

## VI. DISCUSSION AND CONCLUSION

So far, the reported maximum of the (performance-limiting) slowest of the two tolerable polarization fluctuation speeds in two-sided polarization control was 200 rad/s. Our segmented approach with two polarization transformers brings this value to 100 krad/s. This setup also yields roughly 100 times faster tracking than when there is only one polarization transformer. Residual polarization transformations behind polarization transformer 1 are converted into tiny intensity fluctuations. This decouples the two control loops. The two polarization transformers assure a  $2 \times 100$  krad/s tracking on the Poincaré sphere. Insertion loss is 4.7 dB. Permissible optical input power range is at least  $-4 \dots -11$  dBm.

With one polarization transformer, having an insertion loss of 2.2 dB, we have reached  $2 \times (400 \dots 700)$  rad/s tracking, or  $1 \times 1000$  rad/s.

## REFERENCES

- [1] H. C. Lefevre, "Single-mode fiber fractional wave devices and polarisation controllers," *Electron. Lett.* **16**, pp. 778-780, 1980.
- [2] T. Okoshi et al., "New polarization-state control device: Rotatable fiber cranks," *Electron. Lett.* **21**, pp. 895-896, 1985.
- [3] T. Matsumoto, H. Kano, "Endlessly rotatable fractional-wave devices for single-mode-fibre optics," *Electron. Lett.* **22**, pp. 78-79, 1986.
- [4] R. Ulrich, "Polarization stabilization on single-mode fiber," *Appl. Phys. Lett.* **35**, no. 11, pp. 840-842, 1979.
- [5] R. Noe, "Endless polarization control in coherent optical communications," *Electron. Lett.* **22**, pp. 772-773, 1986.
- [6] N. G. Walker, G. R. Walker, "Endless polarization control using four fibre squeezers," *Electron. Lett.* **23**, pp. 290-292, 1987.
- [7] R. Noé et al., "Endless polarization control systems for coherent optics," *IEEE J. Lightwave Techn.* **6**, pp. 1199-1207, 1988.
- [8] N. G. Walker et al., "Endless polarisation control using an integrated optic lithium niobate device", *Electronics Letters* **24** (5), 266-268, 1988.
- [9] B. Koch et al., "100-krad/s Endless Polarisation Tracking with Miniaturised Module Card", *Electronics Letters* **47**, pp. 813-814, 2011.
- [10] B. Koch et al., "Temperature-Insensitive and Two-Sided Endless Polarization Control", *IEEE PTL* **24**, pp. 2077-2079, 2012.

# On Symmetry Properties of Transmission Matrices for Concatenations of Multi Mode Fiber Sections with Linear Coupling

Christian M. Spenner  
Lehrstuhl für Hochfrequenztechnik  
Technische Universität Dortmund  
Dortmund, Germany  
christian.spenner@tu-dortmund.de

Klaus Petermann  
Fachgebiet Hochfrequenztechnik  
Technische Universität Berlin  
Berlin, Germany  
petermann@tu-berlin.de

Peter M. Krummrich  
Lehrstuhl für Hochfrequenztechnik  
Technische Universität Dortmund  
Dortmund, Germany  
peter.krummrich@tu-dortmund.de

**Abstract**—Linear coupling between modes introduces crosstalk in mode division multiplexed transmission systems. We aim to clarify the strength of the contributions of different coupling mechanisms such as fiber axis offset, bending, rotation, etc. to the overall coupling matrix in manufactured fibers. Although the coupling matrices of the individual mechanisms exhibit certain symmetry properties, we find that the overall matrix for a concatenation of multiple fiber sections calculated by the propagation of coherent signals lacks these symmetries. This introduces difficulties for the interpretation of the coupling strength.

Effects such as temperature variations and mechanical vibrations introduce fluctuations of phase differences between propagating modes. These effects result in stochastic interference patterns, which complicates the interpretation of the transmission matrix of a concatenation of elements and potentially break the symmetry. In the incoherent case, the coupling matrix of concatenated elements is more symmetrical due to averaging and changes less over time. We argue that averaged versions of the overall coupling matrices corresponding to the ones for the incoherent case should be used for the evaluation of the coupling strength in a concatenation of multiple fiber sections with discrete coupling points. This facilitates the interpretation of the impact of coupling on system performance.

**Keywords**—fiber optical communication, space division multiplexing, linear mode coupling

## I. INTRODUCTION

Signal transmission in mode division multiplexed (MDM) transmission lines is strongly influenced by signal distortions resulting from linear effects such as differential mode group delay (DMGD) and mode coupling. Analogous to polarization mode dispersion (PMD) and polarization coupling in single-mode fiber (SMF) based communication systems, the signal distortions can be compensated using a multiple-input-multiple-output (MIMO) approach. The complexity of the MIMO equalizer depends strongly on the strength of coupling between modes.

We aim to clarify the amount of coupling that has to be expected in fibers potentially suited for deployment in the field as well as the mechanisms that cause the observed coupling. For this purpose, we try to model the contributions of different coupling mechanisms such as fiber axis offset, bending, rotation, etc. to the overall coupling matrix [1]. These modelled overall coupling matrices are compared with measured coupling matrices of fiber sections.

For this comparison, we need a measure for the coupling strength that is agnostic to temporal fluctuations, which are

caused by temperature changes and varying mechanical stress. The resulting interference patterns change randomly over time and are not meaningful for the evaluation of the general strength of the coupling.

When analyzing calculated overall coupling matrices, we have observed deviations from symmetry, especially for higher order modes. As the individual matrices of the considered discrete coupling points are all symmetric, intuition leads to an expectation of a symmetric overall coupling matrix. We wanted to clarify, whether the deviation from symmetry of the overall matrix potentially originates from numerical problems or the impact of stochastic interference. For this purpose, we investigate the symmetry properties of a fiber link constructed as a concatenation of several sections with discrete coupling points between the sections in the coherent and incoherent case. We believe that our findings are helpful for the evaluation of the coupling strength in multi-mode fiber transmission.

## II. MODELLING OF SIGNAL TRANSMISSION

The signal transmission properties of a linear optical transmission line at an angular frequency  $\omega$  can be described by the typical calculus using scattering parameters:

$$\vec{B}(\omega) = \vec{S}(\omega) * \vec{A}(\omega) \quad (1)$$

The elements of the vector  $\vec{A}$  represent complex envelopes of waves travelling into ports of a device, the elements of the vector  $\vec{B}$  complex envelopes of waves emerging from ports. The squared magnitude of a complex envelope describes the power carried by a given propagating wave.

The square matrix  $\vec{S}$  specifies the scattering of the waves inside a device. A fiber enabling the propagation of  $N$  modes can be modelled as a device with  $2N$  ports as shown in Fig. 1.

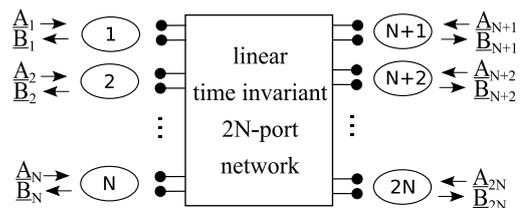


Fig. 1: Modelling of a multi-mode fiber section by scattering parameters

Port 1 corresponds to the fundamental mode at the fiber section input, port  $N+1$  to the same mode at the output, port 2 to the next higher order mode at the input and so on.

In case of fiber based transmission, we are usually interested in signal transmission from the input of the fiber section to its output only. For an easier definition of a respective transmission matrix in forward direction, it is helpful to define a new vector  $\vec{B}'$  of complex envelopes emerging from the ports at the fiber output that contains the coefficients in the lower half of the vector  $\vec{B}$  but with shifted indices:

$$\underline{B}'_n = \underline{B}_{n+N} \text{ for all } n \text{ in the range } 1 \text{ to } N \quad (1)$$

With this new vector and vector  $\vec{A}_{uh}$  containing the coefficient in the upper half of the vector  $\vec{A}$ , a transmission matrix in forward direction  $\vec{T}_F$  with the dimensions  $N \times N$  can be defined as:

$$\vec{B}' = \vec{T}_F * \vec{A}_{uh}. \quad (2)$$

By introducing a vector  $\vec{A}'$  that contains the coefficients in the lower half of the vector  $\vec{A}$  with shifted indices:

$$\underline{A}'_n = \underline{A}_{n+N} \text{ for all } n \text{ in the range } 1 \text{ to } N \quad (3)$$

together with a vector  $\vec{B}_{uh}$  containing the coefficients in the upper half of the vector  $\vec{B}$ , a transmission matrix in backward direction  $\vec{T}_B$  with the dimensions  $N \times N$  similar to the one in forward direction can be defined as

$$\vec{B}_{uh} = \vec{T}_B * \vec{A}'. \quad (4)$$

The coefficients of the matrix  $\vec{T}_F$  correspond to the coefficients of the matrix  $\vec{s}$  in the lower left quadrant and the coefficients of the matrix  $\vec{T}_B$  to the ones in the upper right quadrant.

In case of a reciprocal element, the coefficients of the matrix  $\vec{s}$  are symmetric to the main diagonal, i. e. the matrix is equal to its transposed version:

$$\vec{s} = \vec{s}^T \quad (5)$$

Due to the locations of the coefficients of the transmission matrices in the  $\vec{s}$  matrix, this dictates that the transmission matrix in forward direction has to be equal to the transposed version of the transmission matrix in backward direction:

$$\vec{T}_F = \vec{T}_B^T \quad (6)$$

and vice versa.

For an element with configuration symmetry where the transmission characteristics do not change if the input side is flipped with the output side, the transmission matrices in forward direction and in backward direction have to be equal. A reciprocal element with configuration symmetry results in transmission matrices that are equal and transposed versions of each other. As a consequence, each single matrix features

symmetry of coefficients with respect to the main diagonal. Thus, the transmission matrix of an interface between two fiber sections of equal fiber type that introduces coupling for example due to an offset has to be symmetric.

### III. CONCATENATIONS OF ELEMENTS

In many cases, the transmission matrix of a single element describing discrete coupling is symmetric. For example, this applies to coupling at the interface between two fiber sections of the same fiber type. The overall transmission matrix of the concatenation of two elements corresponds to the product of the individual transmission matrices of the elements. Even if the individual transmission matrices exhibit symmetry, the overall matrix does not have to be symmetric.

In general, the product of two matrices is symmetric if the two matrices commute, i. e. the product is independent of the order of the multiplication:

$$\vec{T}_A * \vec{T}_B = \vec{T}_B * \vec{T}_A. \quad (7)$$

Symmetry of the product only occurs for a very high degree of symmetry of the individual matrices. For example, transmission matrices corresponding to coupling in two mode fiber sections with symmetry to the main diagonal in general do not commute:

$$\begin{aligned} & \begin{bmatrix} t_{1,1} & t_{1,2} \\ t_{1,2} & t_{2,2} \end{bmatrix} * \begin{bmatrix} t'_{1,1} & t'_{1,2} \\ t'_{1,2} & t'_{2,2} \end{bmatrix} \\ & \neq \begin{bmatrix} t'_{1,1} & t'_{1,2} \\ t'_{1,2} & t'_{2,2} \end{bmatrix} * \begin{bmatrix} t_{1,1} & t_{1,2} \\ t_{1,2} & t_{2,2} \end{bmatrix} \end{aligned} \quad (8)$$

They commute, if the individual matrices are symmetric with respect to both diagonals:

$$\vec{T}_{2,DS} = \begin{bmatrix} t_{1,1} & t_{1,2} \\ t_{1,2} & t_{1,1} \end{bmatrix} \quad (9)$$

Two transmission matrices corresponding to coupling in fiber sections with propagation of three modes do not commute, even if the individual matrices feature a very high degree of symmetry:

$$\vec{T}_{3,HS} = \begin{bmatrix} t_{1,1} & t_{1,2} & t_{1,1} \\ t_{1,2} & t_{1,1} & t_{1,2} \\ t_{1,1} & t_{1,2} & t_{1,1} \end{bmatrix} \quad (10)$$

For fiber sections with propagation of four modes, two coupling matrices commute, if the individual matrices feature equal elements on the main diagonal as well as equal elements on the other diagonal and symmetry with respect to both diagonals:

$$\vec{T}_{4,DS} = \begin{bmatrix} t_{1,1} & t_{1,2} & t_{1,3} & t_{1,4} \\ t_{1,2} & t_{1,1} & t_{1,4} & t_{1,3} \\ t_{1,3} & t_{1,4} & t_{1,1} & t_{1,2} \\ t_{1,4} & t_{1,3} & t_{1,2} & t_{1,1} \end{bmatrix} \quad (11)$$

These examples are relevant for coupling in multi mode fibers in cases where coupling occurs only between modes within a given mode group. If the coefficients of the transmission matrix outside the squares which represent the mode groups are equal to zero, the total transmission matrix for a given mode group of a concatenation of two coupling elements corresponds to the product of the individual transmission matrices for this mode group.

#### IV. IMPACT OF INTERFERENCE

As we have seen in the previous section, the overall transmission matrix of a concatenation of two coupling elements in many cases deviates from symmetry, even if the matrices of the individual elements are symmetric. When the product of the two transmission matrices is symmetric, inserting a section with just propagation and no coupling between the two coupling elements can break the symmetry. A Mach-Zehnder-interferometer is analyzed as an example to demonstrate the implications of constructive and destructive interference in the coherent case. A block diagram of the interferometer is shown in Fig. 2.

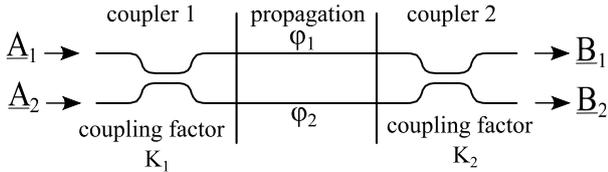


Fig. 2 Block diagram of a Mach-Zehnder-interferometer

The interferometer consists of two couplers with the coupling factors  $K_1$  and  $K_2$  and a section with signal propagation in two branches with unequal optical length in between. The transmission matrices of the couplers and the propagation element can be written as:

$$\overrightarrow{T}_{C1} = \begin{bmatrix} \sqrt{1-K_1} & jK_1 \\ jK_1 & \sqrt{1-K_1} \end{bmatrix}, \quad (12)$$

$$\overrightarrow{T}_{C2} = \begin{bmatrix} \sqrt{1-K_2} & jK_2 \\ jK_2 & \sqrt{1-K_2} \end{bmatrix} \text{ and} \quad (13)$$

$$\overrightarrow{T}_P = \begin{bmatrix} e^{-j\varphi_1} & 0 \\ 0 & e^{-j\varphi_2} \end{bmatrix}. \quad (14)$$

In the matrix  $\overrightarrow{T}_P$  of the propagation element,  $\varphi_1$  and  $\varphi_2$  denote the phase rotations experienced by the signals propagating in the upper and the lower branch. The overall transmission matrix of the interferometer corresponds to the concatenation of the three elements and can be obtained by calculating the product of the individual matrices.

$$\overrightarrow{T}_{MZI} = \begin{bmatrix} t_{1,1} & t_{1,2} \\ t_{2,1} & t_{2,2} \end{bmatrix}. \quad (15)$$

With

$$t_{1,1} = \exp(-j\varphi_1) \sqrt{1-K_1} \sqrt{1-K_2} - \exp(-j\varphi_2) \sqrt{K_1} \sqrt{K_2}, \quad (16)$$

$$t_{1,2} = \exp(-j\varphi_1) \sqrt{K_1} \sqrt{1-K_2} + \exp(-j\varphi_2) \sqrt{1-K_1} \sqrt{K_2}, \quad (17)$$

$$t_{2,1} = \exp(-j\varphi_1) \sqrt{1-K_1} \sqrt{K_2} + \exp(-j\varphi_2) \sqrt{K_1} \sqrt{1-K_2}, \quad (18)$$

$$t_{2,2} = \exp(-j\varphi_2) \sqrt{1-K_1} \sqrt{1-K_2} - \exp(-j\varphi_1) \sqrt{K_1} \sqrt{K_2}, \quad (19)$$

The relation between the off diagonal coefficients  $t_{1,2}$  and  $t_{2,1}$  depends on the phase rotations  $\varphi_1$  and  $\varphi_2$ . In general, the

coefficients are not equal and the overall transmission matrix is not symmetric. The phase rotations are affected by temperature fluctuations and changes of mechanical stress. Stochastic transitions between constructive interference and destructive interference will occur over time. Due to the stochastic fluctuations, a specific case observed at a given point in time does not provide reliable information for the evaluation of the coupling strength.

In the incoherent case, the phase relation between the two signals at the output of the propagation element changes very fast over frequency and over time. Consequently, the time averaged power at a given output of the interferometer does not depend on the phase rotations in the propagation element. A similar independence of the phase rotations is achieved by averaging over all phase differences when calculating the power at a given output. Instead of averaging over the signals at the output, averaging can also be applied to the coefficients of the overall transmission matrix, for example:

$$\overline{|t_{1,1}|^2} = \frac{1}{2\pi} \int_0^{2\pi} |t_{1,1}|^2 d\Delta\varphi = \quad (20)$$

$$(1-K_1)(1-K_2) + K_1K_2$$

with

$$\Delta\varphi = \varphi_2 - \varphi_1. \quad (21)$$

The power transfer matrix of the interferometer  $\overline{PT}_{MZI}$  can be calculated by multiplying the transfer matrices of the individual elements. In this case, only the couplers have to be considered, as the power transfer matrix of the lossless propagation section corresponds to the unit matrix. The power transfer matrices of the two couplers can be written as:

$$\overline{PT}_{C1} = \begin{bmatrix} 1-K_1 & K_1 \\ K_1 & 1-K_1 \end{bmatrix}, \quad (22)$$

$$\overline{PT}_{C2} = \begin{bmatrix} 1-K_2 & K_2 \\ K_2 & 1-K_2 \end{bmatrix}, \quad (23)$$

resulting in the following overall power transfer matrix:

$$\overline{PT}_{MZI} = \begin{bmatrix} (1-K_1)(1-K_2) + K_1K_2 & K_1(1-K_2) + K_2(1-K_1) \\ K_1(1-K_2) + K_2(1-K_1) & (1-K_1)(1-K_2) + K_1K_2 \end{bmatrix}, \quad (24)$$

The coefficients of this matrix are equal to the ones calculated by averaging over the squared magnitudes of the coefficients of the overall transmission matrix for the coherent case. Moreover, the overall power transfer matrix is symmetric. As this matrix does not depend on the phase rotations experienced by the signals in the propagation element and does not fluctuate over time in case of variations of temperature or mechanical stress, it provides a more reliable measure for the evaluation of the coupling strength. Hence, we recommend using the power transfer matrix for the incoherent case for the evaluation of the coupling strength in a concatenation of multi-mode fiber sections with discrete coupling between sections.

## V. CONCATENATIONS OF MANY ELEMENTS

So far, we have investigated the symmetry properties of a single interface between two fiber sections resulting in discrete coupling or the concatenation of two interfaces. In the next step, we want to study the evolution of symmetry properties with increasing number of coupling points.

Defining a measure for the deviation from symmetry facilitates studying the evolution. Therefore, we introduce the measure

$$V_{sym} = \frac{\sum \left| \left| \vec{T}_F \right| - \left| \vec{T}_F^T \right| \right|}{\sum \left| \vec{T}_F \right|} \quad (25)$$

In this equation,  $||$  denotes the element-wise formed magnitude and  $\sum$  the sum of all elements of a matrix.

$V_{sym}$  becomes zero if the matrix  $\vec{T}_F$  is symmetric in magnitude.

## VI. DISCRETE COUPLING MODELL

For simplicity the transmission matrix in forward direction  $\vec{T}_F$  is referred to as  $\vec{T}$  in the following sections. In the discrete coupling model used here,  $\vec{T}$  is composed of a concatenation of  $N$  individual elements, calculated by multiplication of the respective matrices:

$$\vec{T} = \vec{T}_{seg,N} * \dots * \vec{T}_{seg,2} * \vec{T}_{seg,1}. \quad (26)$$

To simplify the analysis of the symmetry of the transmission matrix, a highly simplified model is assumed. Mode coupling occurs only at discrete points at the beginning of each segment due to an offset. No mode coupling takes place inside the fiber sections. In particular, the strong mode coupling that typically occurs within mode groups in gradient fibers is neglected. The model described here is illustrated in Fig. 3.

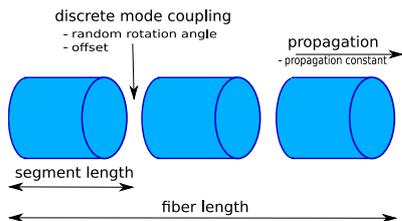


Fig. 3: Illustration of the discrete coupling model

### A. Propagation Matrix

The properties of the transmission matrix are investigated for a graded-index multi-mode fiber supporting 12 true fiber modes, i. e. six modes with two orthogonal polarizations in each one at a wavelength of 1550 nm. The numerically calculated propagation constants  $\beta$  of the propagating modes are shown in Fig. 4.

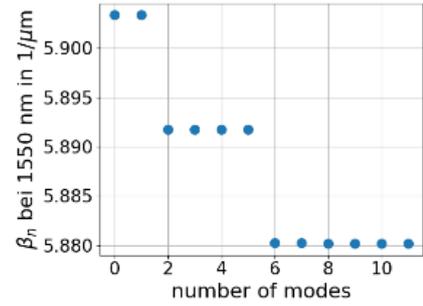


Fig. 4: Propagation constant  $\beta$  of the 12 propagating modes in the investigated fiber at a wavelength of 1550 nm

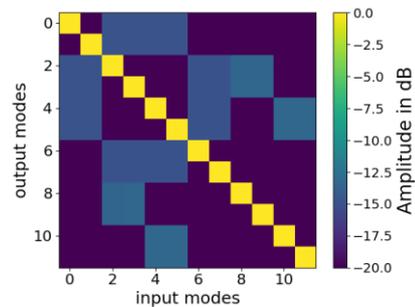
It can be clearly seen that the 12 modes are arranged in three mode groups with very similar propagation constants inside a group. The signal propagation through a fiber section of length  $l$  is taken into account by the diagonal transmission matrix

$$\vec{T}_{prop} = \begin{bmatrix} e^{-j*\beta_1*l} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & e^{-j*\beta_n*l} \end{bmatrix}. \quad (27)$$

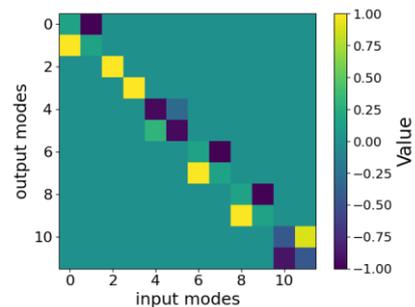
Thereby,  $\beta_n$  represents the propagation constant of the mode  $n$ .

### B. Offset

The mode coupling, which occurs at the discrete offset points, is described by the matrix  $\vec{T}_{offset}$ . It is determined by calculating the overlap integrals between the numerically calculated mode fields for a given offset with an angle of  $45^\circ$  relative to the axis of the Cartesian coordinate system used



(a)



(b)

Fig. 5: Coupling matrix describing an offset of 350 nm with an angle of  $45^\circ$  relative to the orientation of the mode fields (a). Rotation matrix describing a rotation of the mode fields of  $30^\circ$  (b).

for the definition of the mode fields. For an offset between fiber segment axes of 350 nm,  $\vec{T}_{offset}$  is shown in Fig. 5 (a).

It is apparent that power couples mainly into the neighboring mode groups. Power that is not coupled into propagating modes at the offset point results in loss. The matrix  $\vec{T}_{offset}$  is therefore non-unitary. The losses can be interpreted as coupling into radiation modes, which are not captured by the 12 mode transmission matrix used here. Since the strength of coupling into radiation modes increases with mode group order, the losses are mode dependent. Conceivable reflections at the coupling point are neglected in the modeling.

$\vec{T}_{offset}$  is symmetric in magnitude, but the signs are different for the upper and lower triangular sections in the matrix.

### C. Rotation

To avoid having to calculate overlap integrals for  $\vec{T}_{offset}$  for different angles, the matrix  $\vec{T}_{rot}$  is introduced. It transforms the vector with the excitation coefficients into a mode basis, which has the same orientation as the offset. This is realized by multiplication with the unitary matrix  $\vec{T}_{rot}$ . For the true fiber modes used in the simulation,  $\vec{T}_{rot}$  is defined by

$$\begin{bmatrix} \cos(p_1\delta_1) & -\sin(p_1\delta_1) & & & 0 & 0 \\ \sin(p_1\delta_1) & \cos(p_1\delta_1) & \cdots & & 0 & 0 \\ & \vdots & \ddots & & \vdots & \\ & 0 & 0 & \cdots & \cos(p_n\delta_n) & -\sin(p_n\delta_n) \\ & 0 & 0 & & \sin(p_n\delta_n) & \cos(p_n\delta_n) \end{bmatrix} \quad (28)$$

Thereby,  $p_n$  is the azimuthal order of the n-th mode and  $\delta_n$  is the orientation of the offset. In contrast to a LP-Mode representation, where power is exchanged within 2- and 4-folded groups [2], power is just exchanged between pairs of true fiber modes. There are also true fiber modes with an azimuthal order of zero, which are rotationally symmetric. In Fig. 5 (b),  $\vec{T}_{rot}$  is shown for a rotation angle of 30°. The inverse transformation of the excitation coefficients is performed by multiplication with the inverse version of  $\vec{T}_{rot}$ .

## VII. COHERENT SIMULATION

In the coherent case, the input signal is monochromatic and the transmission line is investigated at a given point of time. In the following sections, the symmetry properties of the matrix  $\vec{T}$  are investigated in the coherent case for a single segment, concatenations of identical segments, randomized segments and a symmetric transmission line.

### A. Single Segment

First, the transmission matrix of an individual segment is investigated. A single segment is assumed to consist of a discrete offset point and a segment with a given length. The matrix  $\vec{T}_{seg}$ , which describes the single segment, is therefore calculated by

$$\vec{T}_{seg} = \vec{T}_{prop,seg} * \vec{T}_{rot,seg}^{-1} * \vec{T}_{offset,seg} * \vec{T}_{rot,seg} \quad (29)$$

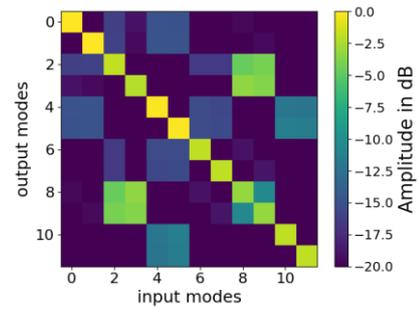
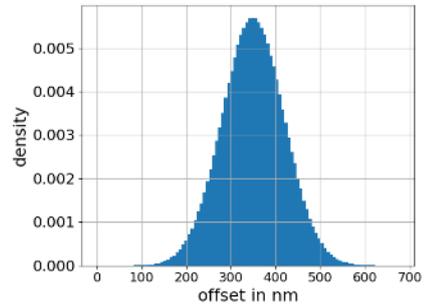


Fig. 6: Overall transmission matrix for a concatenation of 100 identical segments. Each segment has an offset  $l_{offset}$  of 350 nm, an orientation  $\delta$  of 45° and a length  $l_{segment}$  of 100 m.

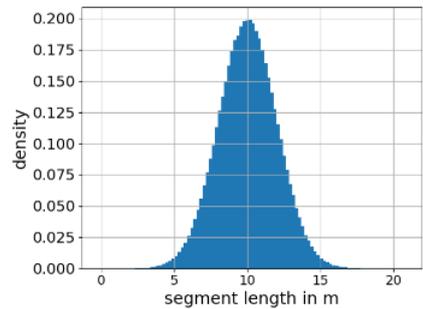
The concatenation of  $\vec{T}_{rot,seg}^{-1} * \vec{T}_{offset,seg} * \vec{T}_{rot,seg}$ , which describes the mode coupling introduced by a single offset point, is symmetric in magnitude. Since the multiplication with the diagonal matrix  $\vec{T}_{prop,seg}$  does not change the symmetry properties of coefficient magnitudes, the transmission matrix of a single segment is also symmetric in magnitude.

### B. Identical Segments

In the special case of a concatenation of  $N$  identical segments, all segment matrices  $\vec{T}_{seg,n}$  are equal and therefore the multiplications commute.



(a)



(b)

Fig. 7 Described distribution of the offsets per segment  $l_{offset}$  (a) and segment length  $l_{segment}$  (b).

Fig. 6 shows the magnitudes of the coefficients in the resulting transfer matrix for a fiber length of 1000 m, 100 segments with a length of 10 m of each one, a rotation angle  $\delta$  of 45° and an offset between fiber axes at the interface

between two segments of 350 nm. The presented matrix is symmetric in magnitude.

### C. Randomized Segments

In the following section, a more realistic transmission case with a sequence of random elements is considered. The offset orientation  $\delta$  for each segment is a random value equally distributed between  $0^\circ$  and  $360^\circ$ . Furthermore, the value of the offset per segment  $l_{offset}$  follows a normal distribution with a mean value of 350 nm and a standard deviation of 70 nm. The corresponding distribution is shown in Fig. 7 (a).

The segment lengths  $l_{segment}$  are also assumed to possess a normal distribution with a mean value of 10 m and a standard deviation of 2 m. This distribution is plotted in Fig. 7 (b). The resulting transfer matrix for a fiber length of 1000 m and the specified distributions for  $\delta$ ,  $l_{offset}$  and  $l_{segment}$  is shown in Fig. 8. It can be seen that the matrix is asymmetric. The calculated value of  $V_{sym}$  for this matrix is 0.12.

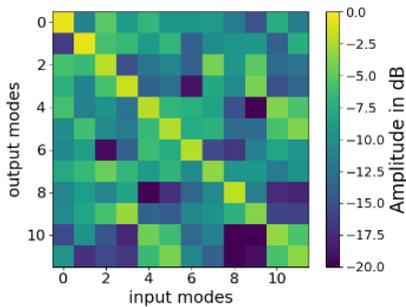


Fig. 8: Overall transmission matrix in the case of randomized segments. From the matrix a value of 0.12 is calculated for  $V_{sym}$ .

### D. Symmetric Transmission Line

For a transmission line with a symmetric configuration, a symmetric transmission matrix is expected. In this case, the  $n$ -th segment from the first half of the transmission line is described analogous to (29) by

$$\vec{T}_n = \vec{T}_{prop,n} * \vec{T}_{rot,n}^{-1} * \vec{T}_{offset,n} * \vec{T}_{rot,n} \quad (30)$$

The considered transmission line is composed of  $N$  segments. The  $n$ -th segment must be equal to the  $(N-n+1)$ -th segment in the symmetric configuration case. The segments in the second half corresponding to the  $n$ -th segment from the first half are described by

$$\vec{T}_{rot,n}^{-1} * \vec{T}_{offset,n}^T * \vec{T}_{rot,n} * \vec{T}_{prop,n} = \vec{T}_{N-n+1} \quad (31)$$

Fig. 9 shows the calculated overall transmission matrix for a transmission line randomly generated according to the described rule.  $\delta$ ,  $l_{offset}$  and  $l_{segment}$  are chosen according to the same distributions as in the previous section. The overall matrix is symmetric in magnitude and phase.

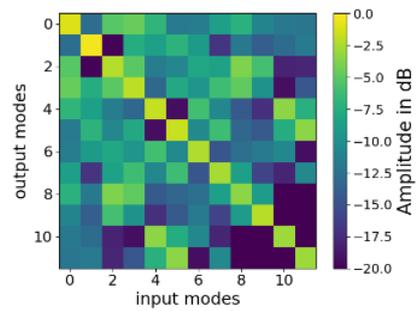


Fig. 9: Overall transmission matrix for a symmetric transmission line. The matrix is symmetrical in magnitude and phase.

## VIII. INCOHERENT SIMULATION

The incoherent case can be realized either by considering a spectrally broad signal or by averaging over a time interval that is significantly longer than the period for which the transmission channel can be assumed to be stationary. Causes for this time variance can be, for example, mechanical or thermal fluctuations.

### A. Model

Here, such fluctuations are modelled by a variation of segment lengths. Due to the different phase constants of the propagating modes, this leads to variable phase conditions at the end of each segment. In the incoherent case, only the magnitude of the transfer matrix can be considered. The matrix  $\vec{T}_{inc}$  for the incoherent case is calculated by

$$\vec{T}_{inc} = \frac{\sum_{k=1}^{K=K} |\vec{T}_k|^2}{K} \quad (32)$$

$\vec{T}_{inc}$  is averaged over  $K$  different coherent matrices  $\vec{T}_k$ . These have the same values for  $\delta$  and  $l_{offset}$ . They differ in a variation of the segment length  $l_{segment}$ . The variation from the original length is given by a normal distribution with a standard deviation of  $\Delta_{l,seg}$ .

### B. Results

In Fig. 10 (a), results are shown from the investigation of the required segment length variation, which is necessary in order to reach the incoherent case. The x-axis specifies the segment length  $\Delta_{l,seg}$ . For each examined value of  $\Delta_{l,seg}$ , 10 random fiber realizations are evaluated. The random values  $\delta$ ,  $l_{seg}$  and  $l_{offset}$  for each segment of these 10 random fiber realizations are chosen from the distributions presented in the previous section. Each realization corresponds to averaging over a number of  $K=100$  segment length variations. The y-axis specifies the value  $V_{sym}$  for the incoherent matrix  $\vec{T}_{inc}$ . For small variations of the segment length,  $V_{sym}$  has a value in the order of 0.11. From a variance of about 0.002 mm, the value of  $V_{sym}$  starts to decrease until a floor is reached at about 0.1 mm.

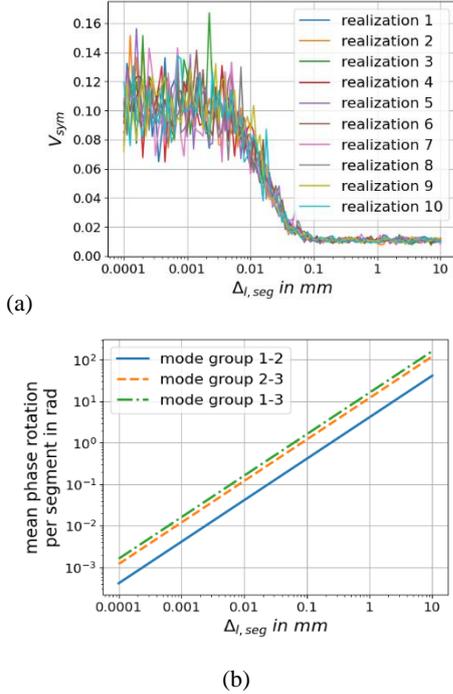


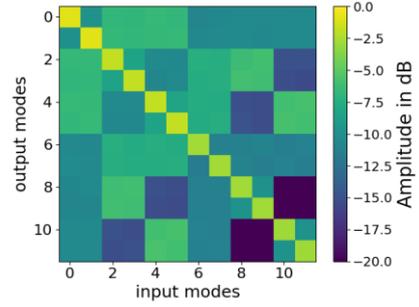
Fig. 10: Symmetry value  $V_{sym}$  versus segment length variance  $\Delta_{l,seg}$  (a). Mean relative phase rotation between the three propagating mode groups (b).

The increase of the symmetry of the matrix coefficient magnitudes is caused by a randomization of the phase conditions between the coupled modes. Since the investigated perturbation resulting from an offset leads primarily to coupling between the mode groups, it can be assumed that the phase conditions between the mode groups are decisive for this behavior. The mean relative phase rotations between the different mode groups are plotted over the segment length variation in Fig. 10 (b). It can be observed that it is necessary for the randomization of the phase relationship between the coupled modes to reach the order of 1 rad at the end of the segments to achieve the incoherent case.

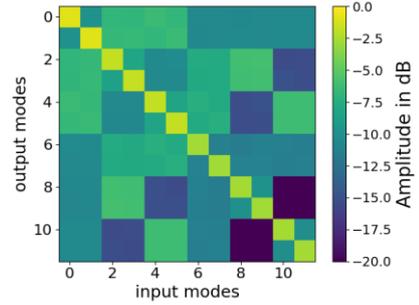
In Fig. 11 two matrices for two different fiber realizations in the incoherent case are shown. It can be seen that both have the same magnitude pattern. From this, it can be concluded that the pattern in the incoherent case does not depend on the exact distribution of offsets, segment lengths and orientations. It should be noted that the pattern shown here is characteristic for the perturbation of offset and bending [3,4].

The results indicate that the deviation from symmetry of the power transfer matrix in the incoherent case can be used as an indication for the effective number of coupling events. A single discrete coupling event tends to result in a symmetric power transfer matrix. The overall power transfer matrices of small numbers of coupling points larger than one tend to deviate from symmetry, whereas a convergence towards more symmetry can be expected for larger numbers of coupling points.

In practice, the strong and frequent coupling within the mode groups will dominate the pattern for the analyzed graded-index multi mode fiber.



(a)



(b)

Fig. 11: Two overall transmission matrices in the incoherent case for different random fiber realizations.

### C. Crosstalk Value Distribution

Due to the usually observed strong mode coupling within the mode groups in manufactured graded-index multi-mode fibers, a characterization of the mode coupling properties is performed by evaluating the coupling between the mode groups. For this purpose, the crosstalk value

$$XT = \frac{\sum |\vec{T}|^2 - \sum |\vec{T}_{main\ diagonal}|^2}{\sum |\vec{T}_{main\ diagonal}|^2} \quad (33)$$

is introduced [5].  $\sum |\vec{T}|^2$  is the overall power in a power transfer matrix. This represents the total power at the output of the fiber when all input modes are excited with the same

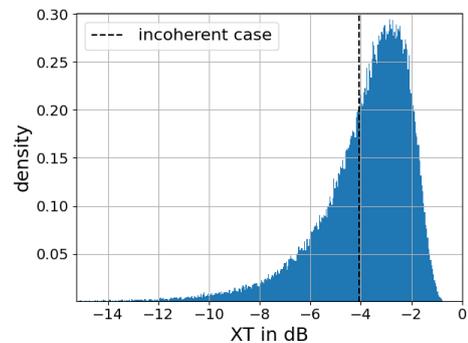


Fig. 12: Distribution of the X-talk-value XT for 100000 randomly generated fibers with each 100 segments and a mean offset per segment of 350 nm.

normalized power.  $\sum \left| \vec{T}_{main\ diagonal} \right|^2$  is the amount of power, which is still detected in the excited mode group. Fig. 12 shows a histogram of the calculated crosstalk values for 100000 randomly generated fiber realizations. The random parameters are generated according to the distribution shown in Section IV. C. For one of the fiber realization, the incoherent case is calculated and presented as the black dashed line.

Analogous to the distribution of differential group delays in PMD, a Maxwell distribution is observed. The individual crosstalk values fluctuate strongly around the average value. The measurement of the crosstalk value for a single coherent state may therefore result in a highly distorted picture of the mode coupling properties of the transmission line. This is the case even though the definition of the crosstalk value itself already averages over the modes within a mode group and over all mode groups. A general statement about the behavior of the mode coupling can only be determined by averaging over many different coherent states. The consideration of the incoherent case corresponds to such an averaging.

#### IX. SUMMARY AND CONCLUSIONS

We have investigated the symmetry properties of transmission matrices describing a concatenation of multi-mode fiber sections with discrete coupling between the sections. Although the individual coupling points are described by a symmetric transmission matrix, concatenations with multiple coupling points typically result in non-symmetric overall transmission matrices in the coherent case. An exception can be found in a transmission line, which consists of identical segments, each one described by a symmetric transmission matrix. Another exception is

given by a transmission line with configuration symmetry, thus featuring the same sequence of differing elements in both directions.

In the incoherent case, the random phase conditions at each coupling point cause the overall matrix to converge to a result with a pattern independent of the specific random realization of the fiber, given that the averaging time has been selected sufficiently long. The pattern does not depend on interference effects impacted by variations of temperature or mechanical stress. The matrix for the incoherent case is therefore better suited for evaluating the strength of mode coupling properties of multi-mode fibers. Moreover, the amount of symmetry of the power transfer matrix can serve as an indication of the number of coupling events.

#### REFERENCES

- [1] C. M. Spenner, P. M. Krummrich and K. Petermann, "Measurement Based Modeling of Mode Coupling Effects in a 50  $\mu\text{m}$  Graded Index Multi Mode Fiber," Photonic Networks; 22th ITG Symposium, 2021, pp. 1-5.
- [2] L. Palmieri, "Coupling mechanism in multimode fibers," Proc. SPIE 9009, Next-Generation Optical Communication: Components, Sub-Systems, and Systems III, 90090G (2014).
- [3] H. Buelow, H. Al-Hashimi, B. T. Abebe and B. Schmauss, "Capacity and Outage of Multimode Fiber with Statistical Bends," 2012 Optical Fiber Communications Conference and Exhibition (OFC), Los Angeles, CA, pp. 1-3 (2012).
- [4] A. A. Juarez, E. Krune, S. Warm, C. A. Bunge, Member and K. Petermann, "Modeling of Mode Coupling in Multimode Fibers With Respect to Bandwidth and Loss," Journal of Lightwave Technologie, Vol. 32, No. 8, pp. 1549-1558 (2014).
- [5] M. Mazur et al., "Characterization of Long Multi-Mode Fiber Links using Digital Holography," 2019 Optical Fiber Communications Conference and Exhibition (OFC), San Diego, CA, USA, 2019, p

# Introducing $\gamma$ -lifting for Learning Nonlinear Pulse Shaping in Coherent Optical Communication

Tim Uhlemann, Alexander Span, Sebastian Dörner, and Stephan ten Brink *Institute of Telecommunications*  
*University of Stuttgart*  
 Pfaffenwaldring 47, 70569 Stuttgart, Germany  
 {uhlemann,span,doerner,tenbrink}@inue.uni-stuttgart.de

**Abstract**—Pulse shaping for coherent optical fiber communication has been an active area of research for the past decade. Most of the early schemes are based on classic Nyquist pulse shaping that was originally intended for linear channels. The best known classic scheme, the *split digital back propagation*, uses joint predistortion and post-equalization and hence, a *nonlinear* transmitter; it, however, suffers from spectral broadening on the fiber due to the Kerr-effect. With the advent of deep learning in communications, it has been realized that an “autoencoder” can learn to communicate efficiently over the optical fiber channel, jointly optimizing geometric constellations and pulse shaping – while also taking into account linear and nonlinear impairments such as chromatic dispersion and Kerr-nonlinearity. E.g., [1] shows how an autoencoder can learn to mitigate spectral broadening due to the Kerr-effect using a trainable *linear* transmitter. In this paper, we extend this *linear* architectural template to a scalable *nonlinear pulse shaping* consisting of a convolutional neural network at both transmitter and receiver. By introducing a novel  $\gamma$ -lifting training procedure tailored to the nonlinear optical fiber channel, we achieve stable autoencoder convergence to pulse shapes reaching information rates outperforming the classic split digital back propagation reference at high input powers.

**Index Terms**—autoencoder, communication, optical, coherent, nonlinear, chromatic dispersion

## 1. Introduction

For the design of an optical fiber communication system that achieves a high spectral efficiency (SE) one requires higher order modulation formats and thus a high signal-to-noise-ratio (SNR) at the receiver (RX). This, in turn, needs higher launch powers at the transmitter (TX). Furthermore, a time-bandwidth-product (TBP) close to one should be chosen to not limit the SE, i.e., symbol rate and occupied bandwidth should almost equal. This, in general, holds for the copper and wireless (radio frequency) channel. In contrast, for the optical fiber, a high input power results in a large Kerr-nonlinearity and hence, in a highly distorted signal at the RX [2]. Apart from the signal-noise interaction, in a single channel scenario, this nonlinearity-induced distortion is deterministic and may be mitigated at TX or RX. Nevertheless, to obtain an optimal equalization performance, one is either forced to take into account the Kerr-nonlinearity-induced spectral

broadening and, thus, support a wider bandwidth compared to the chosen symbol rate [3], [4]; or, operate in a regime where spectral broadening can be neglected [5]. The first approach reduces SE, whereas the second limits design and hardware options; For instance regarding the symbol rate, as spectral broadening inversely increases with it (as will be shown later).

Our superimposed goal is to develop a communication system that is able to overcome the need for additional bandwidth but still achieves optimal (or at least a better) performance than conventional reference systems.

For those, digital signal processing (DSP) algorithms have shown to outperform physical methods such as optical phase conjugation [6], or optical back propagation [7]. Digital approaches can be further divided into three types, (i) a static, analytical, and conventional equalization by fixed algorithms, such as the Volterra-series [8], digital back propagation (DBP) [9], and perturbation models [10]; (ii) nonlinear Fourier transform (NFT)-based schemes [11]; or (iii) deep-learning (DL) techniques with architectural templates based on the aforementioned conventional systems – or based on, e.g., a neural network (NN) [12].

As will be shown later, the DBP is able to compensate for channel impairments almost perfectly up to a certain launch power where signal-noise interaction becomes relevant. This is already known and expected as it is the ideal inverse of the split-step Fourier method (SSFM)-algorithm used to simulate the optical channel. Nevertheless, by limiting the bandwidth of TX and RX the performance decreases significantly due to spectral broadening only. This way we can show that the later introduced autoencoder (AE) is able to compensate for spectral broadening and, hence, allows to operate at higher input powers when compared to conventional systems. One particularity is that training the proposed structure requires a specific procedure that we refer to as “ $\gamma$ -lifting” to achieve stable convergence, and to find better local optima.

This paper, thereby, reports significant progress over [1], where an AE with linear TX was trained to mitigate spectral broadening at high input powers. The training results there have shown that the AE is still not fully able to compensate for the aforementioned nonlinear impairments at higher input powers  $P > 5$  dBm. Hence, here, we extend the AE of [1] to include a nonlinearity compensation (NLC) architecture, effectively implementing a nonlinear TX. Note that the linear TX of [1] is still contained within the new model as a special case, allowing simple evaluation of the additional gains. The

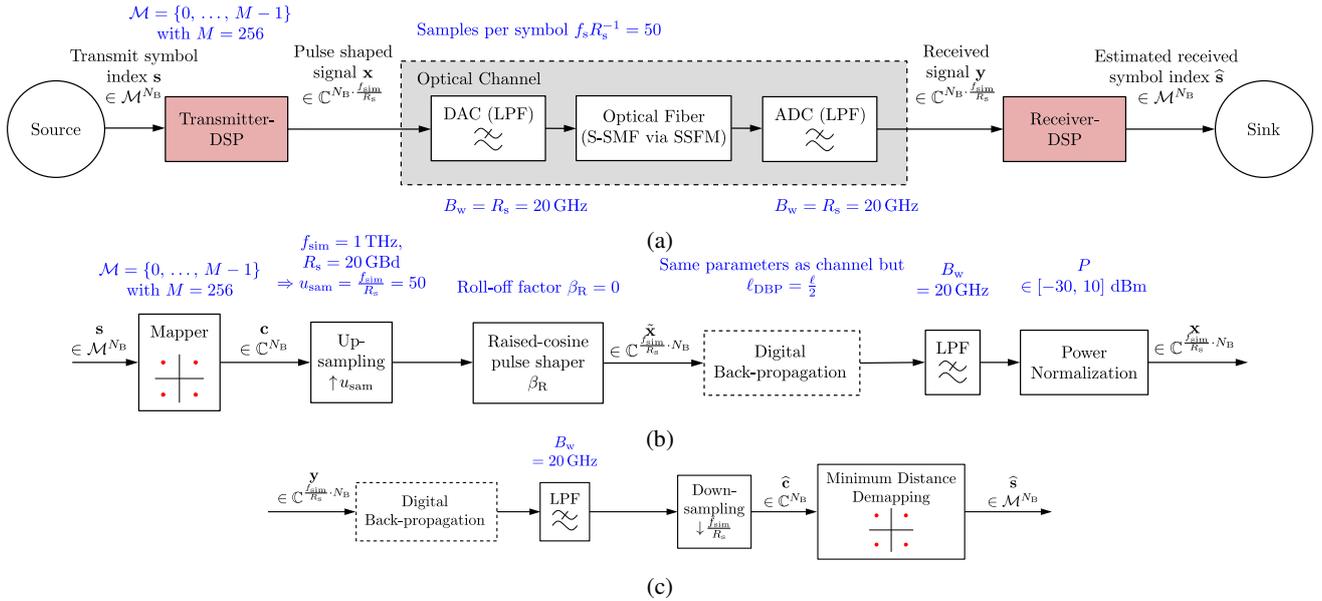


Figure 1: System model with (a) overview that holds for conventional as well as AE setup, (b) the DSP for the conventional TX, and (c) the RX based on split DBP.

nonlinear extension is thereby inspired by the split DBP consisting of a nonlinear and joint mitigation at TX and RX, currently providing the best performance [13].

DL-techniques are already widely used in almost all parts of an optical communication system. The novelty of this paper is the application of an architectural template with nonlinear TX, i.e., its pulse shaping, as well as its particular training method. In a similar spirit, e.g., Häger et al. [14], proposed a trainable SSFM, or DBP, at the RX of which the first half of the compensating path may be moved to the TX and used as predistortion (PD). Despite the fact that our trainable architectural template is not based on the SSFM, we train the PD directly as part of the TX. This allows the TX to learn the opportunities of a PD seeing all subsequent components. Neskorniuk et al. [15] have used a trainable nonlinear PD based on cubic correction terms over an approximated channel model applying the perturbation method. There, they have shown that the training can be significantly stabilized and accelerated up to a specific input power. However, we operate at input powers where this alternative channel model differs too much from the more accurate SSFM, which we have applied in this work. Further, we also included the quasi-analog pulse shaping into the training. Gaiarin et al. [16] have trained a transmitter based on the NFT [17], providing another nonlinear TX structure. Yet, it has been shown that NFT-based systems are currently far from spectrally efficient operations [16].

The remainder of this work is structured as follows. Section 2 provides a short description of the channel, the considered DBP, as well as its performance drop at high launch powers. In section 3 the architectural template of the AE is introduced including the nonlinear TX. Section 4 proposes a new training procedure tailored to the optical fiber channel that leads to improved equalization capabilities of the AE. Finally, section 5 renders some conclusions and provides an outlook to future extensions.

## 2. System model and performance references

The channel model used in this work is taken from [1] comprising the band-limited digital to analog converter (DAC) (represented by a lowpass filter (LPF)), the optical fiber, and a band-limited analog to digital converter (ADC), i.e., again an LPF as shown in the dashed gray box of Fig. 1a. All LPFs are ideally rectangular shaped. This model not only holds for the reference system but also for the AE as introduced later.

The optical fiber is simulated via the symmetric SSFM following the Wiener-Hammerstein model and, thus, approximating a solution of the nonlinear Schrödinger equation (NLSE) [18]

$$\frac{\partial q(t, z)}{\partial z} = j \frac{\beta_2}{2} \frac{\partial^2 q(t, z)}{\partial t^2} - j \gamma |q(t, z)|^2 q(t, z) + n(t, z), \quad (1)$$

where ideal Raman amplification is assumed and the term accounting for fiber attenuation  $\alpha$  was removed. Here  $q(t, z)$  is the optical baseband signal with time  $t$  and distance  $z$ . Alongside mitigating the attenuation, the amplification introduces noise  $n(t, z)$  with power spectral density

$$\rho_n = n_{\text{sp}} h f_0 \alpha. \quad (2)$$

The corresponding channel parameters are summarized in Tab. 1. As we consider a single channel and single polarization system the only observable manifestation of the Kerr-effect is self phase modulation (SPM).

Among the fiber parameters, also some of the signal processing, e.g., those of the converters or DSPs, are fixed to obtain comparable systems; e.g., the supported bandwidth of both converters is set to  $B_w = 20$  GHz. To avoid artificially limiting the SE we have chosen a symbol rate of  $R_s = B_w = 20$  GBd and hence, a symbol duration of  $T = 1/R_s = 50$  ps. The simulation bandwidth is  $f_{\text{sim}} = 1$  THz such that one obtains  $f_{\text{sim}} R_s^{-1} = 50$  samples per symbol. The resulting TBP is

$$\Pi_{\text{TBP}} = T \cdot B_w = 1. \quad (3)$$

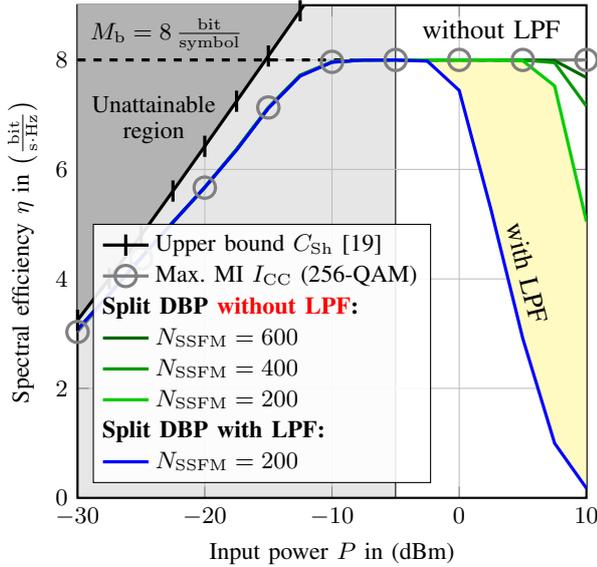


Figure 2: Comparison of the performance of a split DBP with and without (DAC/ADC-induced) LPFs.

The alphabet, or set of message symbol indices, is  $\mathcal{M} = \{0, \dots, M-1\}$  with  $M = 256$  and the number of transmitted symbols is  $N_B$ .

As performance measure, we use the spectral efficiency (SE) defined as

$$\eta = \frac{I(s; \hat{s})}{T \cdot B_w} = \frac{I(s; \hat{s})}{\Pi_{TB}} \stackrel{\Pi_{TB}=1}{=} I(s; \hat{s}), \quad (4)$$

where  $I(s; \hat{s})$  is the mutual information (MI) between the distributions of the transmitted  $s$  and the received (hard decided) symbol indices  $\hat{s}$  as shown in Fig. 1a. Here we use histograms of hard decided symbols at RX to include the whole RX-DSP in the chosen performance measure. Obviously, the performance could be improved by using soft outputs. But, in this work, we are more interested in studying how to mitigate the drop of MI at high launch power and how to address this by means of pulse shaping.

TABLE 1: Simulation parameters.

Property	Symbol	Value
Planck's constant	$h$	$6.626 \cdot 10^{-34}$ J s
Carrier frequency	$f_0$	193.55 THz
Attenuation	$\alpha$	$0.046 \text{ km}^{-1}$ $\cong 0.2 \text{ dB km}^{-1}$
Chromatic dispersion	$\beta_2$	$-21.67 \text{ ps}^2 \text{ km}^{-1}$
Kerr-nonlinearity	$\gamma$	$1.27 \text{ km}^{-1} \text{ W}^{-1}$
Spontaneous emission	$n_{sp}$	1
Fiber length	$\ell$	1000 km
Simulation sampling rate	$f_{sim}$	1 THz
No. of SSFM-steps	$N_{SSFM}$	200
Fixed SSFM step-size	$\Delta z_{SSFM}$	5 km
Launch power	$P$	-30 dBm, ..., 10 dBm
Bandwidth of TX/RX	$B_w$	20 GHz

## 2.1. Spectral broadening revisited

Taking into account the considered power range for  $P$ , we obtain a dispersion length of

$$\ell_D = T^2 / |\beta_2| = 115 \text{ km} \quad (5)$$

and a nonlinear length of

$$\ell_{NL}(P) = 1/(\gamma P) \in [78, 78 \cdot 10^4] \text{ km} \quad (6)$$

such that  $\ell_D < \ell_{NL}(P)$  does not hold for all input powers  $P$  [18]. This way we operate in a regime where spectral broadening cannot be neglected [20]. Further it holds that

$$\ell_{NL}(P) < \ell \forall P > -1.0 \text{ dBm}. \quad (7)$$

Consequently, the received signal's bandwidth definitely exceeds the supported system bandwidth  $B_w = R_s$  for conventional communication systems (and their corresponding equalization) as there is no margin implemented between  $B_w$  and the original signal bandwidth  $R_s$ .<sup>1</sup> Next, we specify the goal for the AE more precisely to "Learning a high MI despite the effect of spectral broadening."

## 2.2. Limitations of Digital Back Propagation

As a reference system we use the "split DBP" or also called "split NLC" by [13]. The corresponding TX and RX are shown in Fig. 1b and Fig. 1c, respectively.

The TX consists of a conventional 256-quadrature-amplitude-modulation (QAM) that translates the symbol indices  $\mathbf{s} \in \mathcal{M}^{N_B}$  to complex constellation symbols  $\mathbf{c} \in \mathbb{C}^{N_B}$ . An ideal Nyquist pulse-shaping (consisting of an upsampling to simulation frequency  $f_{sim}$  and a raised cosine pulse shaping with roll-off  $\beta_R = 0$ ) lead to the bare transmit signal  $\tilde{\mathbf{x}}$ . Typically a two-fold oversampling is used, whereas we use an oversampling of  $u_{sam} = f_{sim}/R_s = 50$  for a fair comparison with the later introduced trainable models that shall not learn to compensate any inaccuracy, and to exclude a numerical mismatch between channel and equalizer. It follows an NLC as PD that is realized by a DBP, which pre-compensates for a fiber length of  $\ell_{DBP} = \frac{\ell}{2}$ . A subsequent LPF plus power normalization generates the transmit signal  $\mathbf{x}$ . The RX starts with a DBP processing the received signal  $\mathbf{y}$  for a fiber length of again  $\ell_{DBP} = \frac{\ell}{2}$  at simulation rate, which is already lowpass-filtered by the channel (see converters in Fig. 1a).<sup>2</sup> An LPF assures a correct down-sampling to symbol rate  $R_s$  for a subsequent minimum distance demapping. The DBP algorithm itself is based on the symmetric SSFM following the Wiener-Hammerstein model as shown in [2] that is also used for the channel simulation.

It was already shown that for a PD-only implementation "SPM can be eliminated, within experimental error, up to where the transmitter bandwidth limits are reached" [21]. Hence, in general the chosen PD leads to a transmit signal that perfectly pre-compensates for channel impairments but occupies more bandwidth than the supported one by the converters. The same holds for the

1. Note that a margin would artificially increase the TBP and, hence, lower the SE.

2. In the simulation the ADC belongs to the channel, but technically the channel has no lowpass-characteristic.

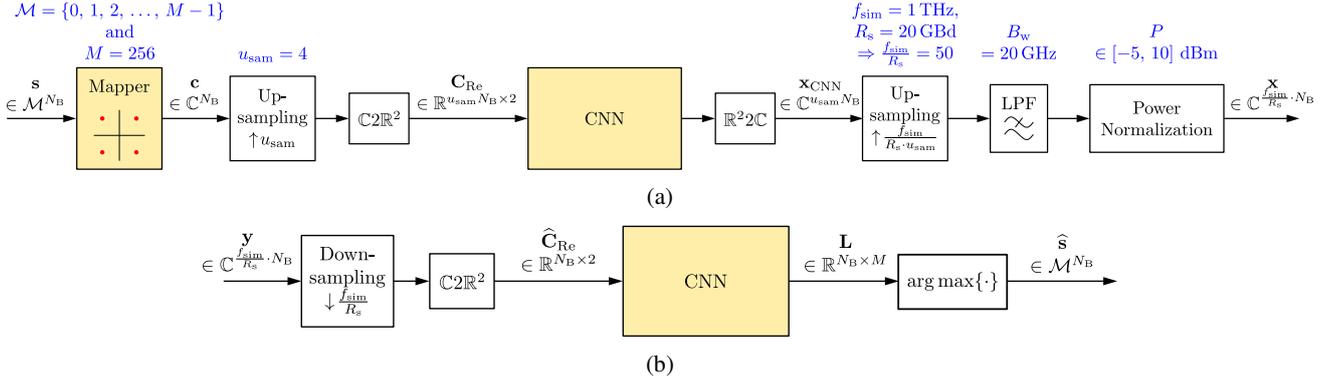


Figure 3: Architectural template of the proposed AE implementing the TX- and RX-DSP. Trainable blocks are colored in yellow. (a) shows the TX-, and (b) the RX-DSP as introduced in this work.

application of a post-equalization (PE)-only where now the receiver bandwidth limits the resulting performance. Hence, to overcome spectral broadening introduced by Kerr-nonlinearity, one has to take into account an artificial bandwidth expansion. This required expansion combined with the LPF explains the performance drop of the split DBP at high launch powers.

The curves of Fig. 2 confirms the statement above. In green it shows the SE of the split DBP without the aforementioned LPFs. Indeed, it achieves the expected near optimal performance up to a certain launch power, where the drop can be related to, either, the limited accuracy of the SSFM, i.e., the number of steps  $N_{\text{SSFM}}$  used for the channel and the DBP, or, signal-noise-interaction. To get a better understanding of the accuracy, this system is evaluated at different number of SSFM steps. Enabling the LPFs one obtains the blue curve exhibiting a drop at lower launch powers than before. The space in between green and blue curve depicts the loss introduced only by spectral broadening.

For comparison (i) an upper bound proposed by Kramer et al. [19] that is equivalent to the Shannon limit and (ii) the constellation constrained MI  $I_{CC}$  for an additive white Gaussian noise (AWGN) channel based on hard decisions and assuming a 256-QAM are shown. The latter curve only indicates that, for low input powers, the conventional system achieves an almost optimal equalization. For high input powers,  $I_{CC}$  saturates as expected due to the constrained constellation.

### 2.3. Discussion of Conventional Reference Curves

Even if the DBP currently shows the best performance, and indeed compensates for the nonlinear impairments almost perfectly, this only holds for the case where, either, spectral broadening is not relevant, or, DAC, and/or ADC support a larger bandwidth than required by the actual information, at the expense of a higher TBP and, hence, lower SE by design. This gives rise to the assumption that better solutions than DBP in terms of SE exist for the case where both converters significantly limit the bandwidth. I.e., the goal is to enter the attainable region between the curves with and without LPF, indicated in yellow in Fig. 2. While this goal has not changed since [1], we now can apply both an extended AE and a new training procedure to further close this gap.

## 3. Autoencoder-optimized pulse shaping

Fig. 3 shows the block diagram of the new architectural template for the AE consisting of the TX in Fig. 3a and RX in Fig. 3b, where now both are trainable and nonlinear.

### 3.1. Brief recap of neural network notation

A network consists of an ordered set of layers  $\mathcal{L} = \{L_1, L_2, \dots, L_\Theta\}$ , where  $L_\vartheta$  is the  $\vartheta$ th layer of the total number of  $\Theta$  layers. Here, a layer is

$$L_\vartheta : \mathbb{R}^{K_{\vartheta-1} \times D_{\vartheta-1}} \rightarrow \mathbb{R}^{K_\vartheta \times D_\vartheta} \quad (8)$$

$$\mathbf{U}_{\vartheta-1} \mapsto \mathbf{U}_\vartheta$$

where  $\mathbf{U}_{\vartheta-1}$  is the input and  $\mathbf{U}_\vartheta$  is the output signal of the  $\vartheta$ th layer;  $K_\vartheta \in \mathbb{N}$  is the length of the signal in terms of time samples after the  $\vartheta$ th layer; and  $D_\vartheta$  is the dimensionality of each time sample, i.e., the number of channel components. An example of channel components are real and imaginary part of a signal with  $D_\vartheta = 2$ . In the following  $u_{\vartheta,k,d} = [\mathbf{U}_\vartheta]_{k,d}$  is element of a signal matrix at time instance  $k$  and component  $d$ . In this work a layer is either a convolutional or a dense layer, where  $\vartheta = 0$  denotes the input signal.

For a convolutional layer  $L_\vartheta = L_{\vartheta,\text{conv}}$ , the output dimension  $D_\vartheta$  corresponds to the number of different trainable kernels or filters  $\mathbf{H}_{\vartheta,d} \in \mathbb{R}^{N_\vartheta \times D_{\vartheta-1}}$  for  $d = 0, \dots, D_\vartheta - 1$ ;  $N_\vartheta < K_\vartheta$  is the filter's individual length. All filters of a single layer have the same length. The output sample of a convolutional layer  $\vartheta$  at time instance  $k$  and for output component  $d$  can be formulated as

$$u_{\vartheta,k,d} = L_{\vartheta,\text{conv}}(\mathbf{U}_{\vartheta-1}) = \varphi_\vartheta \left\{ \text{tr} \left( \bar{\mathbf{U}}_{\vartheta-1,k} \cdot \mathbf{H}_{\vartheta,d}^T \right) + b_{\vartheta,d} \right\} \quad (9)$$

where  $\varphi_\vartheta$  is the activation function,  $(\cdot)^T$  is the transpose,  $b_{\vartheta,d}$  is a trainable bias, and

$$\bar{\mathbf{U}}_{\vartheta-1,k} = \begin{pmatrix} u_{\vartheta-1,k-\bar{N},0} & \cdots & u_{\vartheta-1,k-\bar{N},D_{\vartheta-1}-1} \\ \vdots & \ddots & \vdots \\ u_{\vartheta-1,k+\bar{N},0} & \cdots & u_{\vartheta-1,k+\bar{N},D_{\vartheta-1}-1} \end{pmatrix} \quad (10)$$

is a sliced and shifted version of the input signal with  $\bar{N} = (N_\vartheta - 1)/2$ . The filter length  $N_\vartheta$  is always chosen

to be odd, to have a symmetric influence of the neighboring samples. A zero padding is added such that one obtains equal length of the input and output signal what is often referred to as “same” padding. This corresponds to  $u_{\vartheta,k,d} = 0 \forall k < 0$  or  $k \geq K_{\vartheta}$ .

For a dense layer  $L_{\vartheta} = L_{\vartheta,\text{dense}}$ , the output dimension  $D_{\vartheta}$  corresponds to the number of neurons of that layer and  $N_{\vartheta} = 1 \forall \vartheta$  as, here, each layer operates on a single time instance  $k$ . This simplifies Eq. (9) to

$$\begin{aligned} u_{\vartheta,k,d} &= L_{\vartheta,\text{dense}}(\mathbf{U}_{\vartheta-1}) \\ &= \varphi_{\vartheta} \{ [\mathbf{U}_{\vartheta-1}]_k \cdot \mathbf{H}_{\vartheta,d}^T + b_{\vartheta,d} \}. \end{aligned} \quad (11)$$

### 3.2. Transmitter Design

Equivalent to the reference system, the symbol indices  $\mathbf{s} \in \mathcal{M}^{N_B}$  shall be transmitted by the TX and estimated by the RX with a symbol rate of  $R_s = 20$  GBd. Here, the same system parameters hold as already given in Sec. 2. Each transmission contains a batch of  $N_B = 12,000$  symbol indices, which is chosen on the basis of available GPU-memory in our simulation setup.

The TX as given in Fig. 3a starts with a trainable mapping, i.e., an  $M \times 2$ -matrix, choosing symbols  $\mathbf{c}$  from a complex constellation based on the incoming symbol indices  $\mathbf{s}$ . Then, the signal is upsampled from symbol rate  $R_s$  to  $u_{\text{sam}} \cdot R_s$  by adding zeroes in between the in-phase quadrature (IQ)-symbols, and, subsequently, pulse shaped. Here,  $u_{\text{sam}} = 4$  has shown to be sufficiently large to achieve the results in the following figures, staying well below the upsampling of the DBP, where  $u_{\text{sam}} = 50$ .

The nonlinear pulse shaping generates the waveform  $\mathbf{x}_{\text{CNN}}$  and is realized by a convolutional neural network (CNN). As such, it can only process real valued signals, and, thus, the two transformations  $\mathbb{C}2\mathbb{R}^2$  and  $\mathbb{R}^22\mathbb{C}$  vice-versa are required. Thereby, real and imaginary part of the complex signal are stacked in an additional last dimension.

It follows an upsampling from sample rate  $u_{\text{sam}} \cdot R_s = 80$  GHz of the CNN to  $f_{\text{sim}} = 1$  THz, now including a Whittaker-Shannon interpolation for quasi-continuous waveform channel simulation. The lowpass-filtered and power normalized signal  $\mathbf{x}$  is finally transmitted over

TABLE 2: Parameters of the TX- and RX-NN.

Layer	Type	Parameters		
		$D_{\vartheta}$	$N_{\vartheta}$	$\varphi_{\vartheta}$
Transmitter				
1	Convolutional	128	221	linear
2	Dense	64	-	elu
3	Dense	64	-	elu
4	Dense	64	-	elu
5	Dense	2	-	linear
Receiver				
1	Convolutional	128	55	linear
2	Dense	2048	-	elu
3	Dense	2048	-	elu
4	Dense	512	-	elu
5	Dense	512	-	elu
6	Dense	512	-	elu
7	Dense	256	-	linear

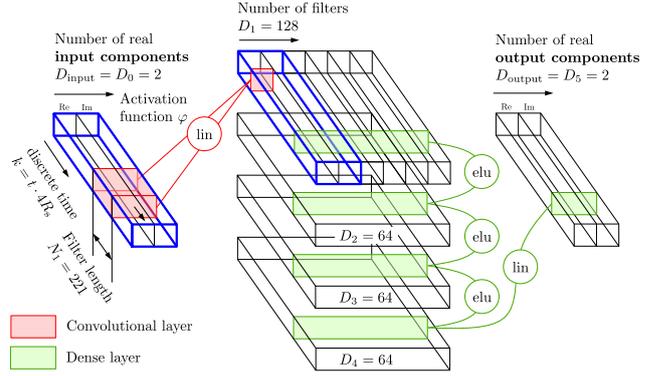


Figure 4: CNN used at TX. Boxes depict the signals  $\mathbf{U}_{\vartheta}$ . Bold blue lines depict the equivalent (linear) structure of [1] with  $D_{\text{output}} = D_1 = 2$ .

the channel, with samples being calculated as  $[\mathbf{x}]_k = g_{\text{TX}}(k, \mathbf{s}; \Omega_{\text{TX}})$ . Here,  $g_{\text{TX}}$  summarizes the TX function with trainable parameters  $\Omega_{\text{TX}}$ , i.e., the weights  $\mathbf{H}_{\vartheta}$ .

The design of the TX CNN of Fig. 3a is shown in Fig. 4, with the parameters given in Tab. 2. It is motivated by the fact that neighboring symbols nonlinearly interact with each other along the fiber. The input signal of the CNN is  $\mathbf{U}_0 = \mathbf{C}_{\text{Re}} \in \mathbb{R}^{u_{\text{sam}} N_B \times 2}$ . Hence, the input length  $K_0 = u_{\text{sam}} N_B$  is the sequence of constellation symbols upsampled by  $u_{\text{sam}}$ . The input dimension  $D_0 = 2$  represents real and imaginary part of the original complex signal. The first layer is a convolutional layer that convolves its  $D_1 = 128$  different filter kernels  $\mathbf{H}_{1,d} \in \mathbb{R}^{N_1 \times D_0}$  for  $d < D_1$  with the real and imaginary part of the input signal. The length

$$N_1 = \lceil \Delta t_{\text{CD}} \cdot R_s \rceil \cdot u_{\text{sam}} + 1 = 221 \quad (12)$$

is chosen such that the filter length is in the order of the pulse widening induced by chromatic dispersion (CD) of

$$\Delta t_{\text{CD}} = 2\pi\beta_2 \ell B_w. \quad (13)$$

Also the dimension  $D_1$  is derived from the pulse widening such that this layer is *at least* able to shift all relevant neighboring symbols to the same time instance by a (probably learned) time shifting filter  $\mathbf{H}_{\text{shift}} = (\delta_{-D_1,k}, \dots, \delta_{D_1,k})$  with Kronecker-Delta

$$\delta_{n,k} = \begin{cases} 1 & k = n \\ 0 & \text{else} \end{cases} \quad (14)$$

and  $k \in \{0, \dots, K_1 - 1\}$ . Hence, we have chosen  $D_1 = 128 \gg \lceil \Delta t_{\text{CD}} \cdot R_s \rceil = 55$ . It follows a sub-network consisting of dense layers that operate on the single time instances of the  $D_1$  convolved signals. The last layer again produces  $D_5 = D_0 = 2$  time signals that form the real and imaginary part of the pre-distorted signal  $\mathbf{x}_{\text{CNN}} = \mathbb{R}^22\mathbb{C}\{\mathbf{U}_5\}$ .

Bold blue lines mark the TX-structure of the former AE applied in [1]. Hence, the case  $D_1 = 2$  and an imposed constraint on the kernel to perform a complex convolution, corresponds to the earlier TX, i.e., the former structure is a special case of the one provided here, respectively.

### 3.3. Receiver Design

The computational efficiency of the overlap & save mechanism from [1] was not sufficient to be exhaustively trained such that we have had to deviate more from this intuitive design to reduce complexity. Hence, what follows is a CNN that substitutes the overlap & save mechanism in a more efficient way while having at least the same degree of freedom. Nevertheless, the achieved gains of the new architecture stem from the TX.

The RX as shown in Fig. 3b consists of blocks for downsampling the receive signal  $\mathbf{y}$  from simulation frequency  $f_{\text{sim}}$  to symbol rate  $R_s$ , a complex-to-real-transformation, and a CNN to learn an estimate of the desired likelihoods  $\hat{l}_{k,m} = \hat{P}(S_k = m|\mathbf{y})$  of the transmitted symbol indices  $S_k$  based on the received signal  $\mathbf{y}$ . Note that  $S_k$  denotes the random variable (RV) corresponding to the true transmitted symbol index  $s_k$  at time instance  $k$  and a realization  $m \in \mathcal{M}$ . This can be obtained by a row-wise softmax-normalization of the CNN's output  $[\mathbf{L}]_{k,m} = \hat{l}_{k,m}$  and appropriate cost function as described below. The whole block can be written as a trainable function such that  $\hat{l}_{k,m} = g_{\text{RX}}(k, m, \mathbf{y}; \Omega_{\text{RX}})$  with parameters  $\Omega_{\text{RX}}$  that consists of the weights  $\mathbf{H}_\theta$  of the RX-NN. In a last step, the symbol with highest likelihood is chosen to be the hard decided estimate

$$\hat{s}_k = \arg \max_m g_{\text{RX}}(k, m, \mathbf{y}; \Omega_{\text{RX}}) \in \mathcal{M}. \quad (15)$$

The CNN has the same architecture as in Fig. 3 but with different parameters, given in Tab. 2. The number of output neurons of the last layer is  $M$  and has a linear activation function. The softmax-normalized output, then, can be interpreted as the previously introduced likelihoods of the transmit symbol.

## 4. Training using $\gamma$ -Lifting

The general objective of the trainable AE is an optimal estimate of the transmitted symbols at RX

$$\hat{\mathbf{s}} \stackrel{!}{=} \mathbf{s} \Rightarrow \hat{s}_k \stackrel{!}{=} s_k \forall k, \quad (16)$$

where  $\mathbf{s}$  is the vector containing the transmitted symbol indices, and  $\hat{\mathbf{s}}$  the corresponding estimate by the RX.

To achieve this goal, the parameters  $\Omega = \{\Omega_{\text{TX}}, \Omega_{\text{RX}}\}$  of TX and RX are trainable, which is common for all AE-setups in communications. Hence, not only TX and RX, but the whole AE system of Fig. 1a can be seen as a trainable function  $g(k, m, \mathbf{s}; \Omega)$ . This way, the estimate of the symbol at time instance  $k$  is the argument of the maximization of the learned probabilities over all symbol indices  $\hat{s}_k = \arg \max_m g(k, m, \mathbf{s}; \Omega)$  with  $m \in \mathcal{M}$ . Having said this, the objective can be further detailed to  $g(k, m, \mathbf{s}; \Omega) = \hat{P}(S_k = m) \stackrel{!}{=} P(S_k = m) \forall k$ , where

$$P(S_k = m) = \begin{cases} 1, & m = s_k \\ 0, & \text{else} \end{cases} \quad (17)$$

is the probability mass function (PMF) of the true transmit symbol indices  $s_k = [\mathbf{s}]_k$  (this refers to ‘‘one-hot encoding’’, as the PMF of the transmitted symbols corresponds to the Kronecker-delta). The optical channel is contained

in  $g(k, m, \mathbf{s}; \Omega)$  and serves as the penalty to overcome by choosing  $\Omega$  appropriately.

Equivalent to [1], we trained the AE using samplewise cross-entropy (CE) loss for batch sample  $k$  defined as

$$\begin{aligned} \mathcal{L}_{\text{CE}} \left( P(S_k = m), \hat{P}(S_k = m) \right) \\ = \mathbb{H}_{P(S_k = m)} \left[ \hat{P}(S_k = m) \right] \end{aligned} \quad (18)$$

where  $\mathbb{H}[\cdot]$  is cross-entropy and  $\mathbb{E}[\cdot]$  is expectation. Taking into account Eq. (17) this can be further simplified to

$$\mathcal{L}_{\text{CE}} \left( P, \hat{P} \right) = -\log g(k, m = s_k, \mathbf{s}; \Omega) \quad (19)$$

The resulting cost is defined over the batch of symbols with size  $N_B$  and calculates as

$$\begin{aligned} \mathcal{C}_{\text{CE}} \left( P(S_k = m), \hat{P}(S_k = m) \right) \\ = -\frac{1}{N_B} \sum_{k=0}^{N_B-1} \log g(k, m = s_k, \mathbf{s}; \Omega) \end{aligned} \quad (20)$$

For the optimizer we have chosen Adam [22], which applies stochastic gradient descent (SGD) to minimize the cost as

$$\Omega_{\text{opt}} = \arg \min_{\Omega} \mathcal{C}_{\text{CE}} \left( P, \hat{P} \right) \quad (21)$$

by training the parameters  $\Omega$ . For a conventional training as in [1] the channel is fixed and only training parameters, i.e., the learning rate  $w_{\text{lr}}$  and batch size  $N_B$ , are adjusted during training to support convergence. The latter is achieved by the aid of the training configuration of Tab. 3.

Nevertheless, the very first results have shown that the training procedure of [1] is not sufficient to train the proposed architectural template. Only by repeating the training given in Tab. 3, by chance, we obtain the dashed red results of Fig. 5b. The underlying problem is depicted also in dashed red in Fig. 5a, where one can see that the performance jumps during the first training cycles, i.e., repetitions. Having realized that convergence is a major problem of training the AE through the SSFM, we further investigated on how to increase the stability of the convergence and to probably find better local optima. Hence, in the following we introduce a new training method that was inspired by the training described in [23]. There, the input training samples were chosen such that the learning gets simplified or, i.e., the relation between input and output becomes more obvious to the trainable structure.

We have searched for a way to simplify the stated problem such that we can gradually converge until finally arriving at the desired more complex problem to solve.

TABLE 3: Configuration of a single training cycle.

Max. training iterations	Batch size	Learning rate
$N_{\text{train}}$	$N_B$	$w_{\text{lr}}$
2000	12,000	$2 \cdot 10^{-4}$
1500	12,000	$1 \cdot 10^{-4}$
1000	12,000	$5 \cdot 10^{-5}$
500	12,000	$1 \cdot 10^{-5}$

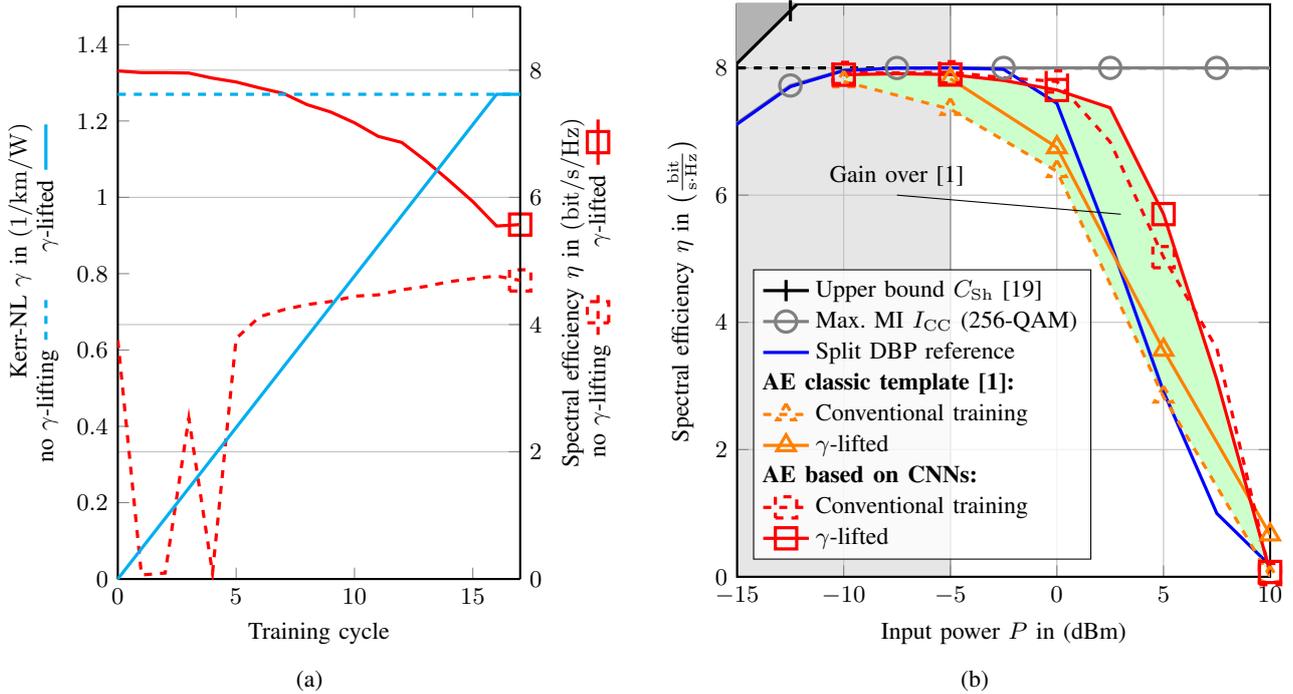


Figure 5: Performance evaluation of the AE based on CNN with (a) showing the  $\gamma$ -lifting that describes the increase of  $\gamma$  over training cycle on the left axis and the resulting SE on the right axis at  $P = 5$  dBm; and (b) over input power compared to the different AEs with  $\gamma$ -lifting (solid) and without (dashed).

Here, this may be achieved by setting some channel parameters to zero and increasing them incrementally, which, in effect, results in an *increasing difficulty* for training the system. Thereby, we ensure that training converges in each cycle, before making it harder. As potential parameters we identified the following:

Slowly increasing fiber length  $\ell$  simplifies the problem significantly as all impairments increase with fiber length. On the downside, it poses an unsolvable conflict between a varying accuracy and a vanishing gradient information. This can be seen by considering the step size of the SSFM  $\Delta z_{\text{SSFM}} = \ell/N_{\text{SSFM}}$ . A growing  $\ell$  means either an increasing  $\Delta z_{\text{SSFM}}$  and hence a varying accuracy; or adaption of  $N_{\text{SSFM}}$ , which means that the error has to be backpropagated through an increasing number of SSFM-iterations. This way, the very first training cycles may be predominant compared to the later ones with weaker gradient information. Using chromatic dispersion  $\beta_2$  means reducing inter-symbol interference (ISI) at the beginning and incrementally adding more influence of neighboring symbols. This allows to start with a simple sample-wise compensation at the beginning, but is in contrast to the requirement that the filter length  $N_\theta$  of the CNN has to be fixed to the final length over all trainings, anyway. Hence, for the first training cycles, most of the randomly initialized trainable filter taps rather distort than may ever support. For considering Kerr-effect only,  $\gamma$  controls the power-dependent rotation of a sample, where, in turn, its power depends on the corresponding transmit symbol, neighboring transmit symbols, and noise. Reducing  $\gamma$  disentangles the simple linear superposition of CD and may help to learn their interactions step-by-step.

As there is no obvious drawback connected to the Kerr-nonlinearity parameter, we decided to use a  $\gamma$ -lifting.

This  $\gamma$ -over-cycle dependency is also depicted in Fig. 5a for  $P = 5$  dBm by the solid cyan curve, which linearly increases with training cycle. Note that, each training cycle consists of all training iterations described by Tab. 3. The conventional training is depicted by the dashed cyan curve showing that  $\gamma$  is constant. It may be misleading that the SE worsens over training cycle, which is a consequence of the increasing difficulty. Nevertheless, comparing both red curves – the one without  $\gamma$ -lifting (dashed) and the one with (solid) – shows that the latter exhibits a significantly stabilized convergence, as no large variations can be observed anymore. Further, it achieved a significantly higher SE for the final  $\gamma = 1.27 \text{ km}^{-1} \text{ W}^{-1}$ , i.e.,  $5.7 \text{ bit}/\text{s}/\text{Hz}$  vs.  $5.0 \text{ bit}/\text{s}/\text{Hz}$ .

Performing  $\gamma$ -lifting over the whole input power space results in Fig. 5b. The dashed curves again show the earlier AE (orange) and the one proposed in this work (red) with their conventional training. Switching on  $\gamma$ -lifting leads to the corresponding solid curves. It can be seen that  $\gamma$ -lifting, in most cases, pushes the performance of the CNN-based AE closer to  $I_{\text{CC}}$ . By chance, thereby, the conventional training may still exceed the  $\gamma$ -lifting due to the high fluctuations in the training quality. Nevertheless, the proposed AE outperforms the former AE as well as the split DBP at high input powers. This means that we have compensated or avoided spectral broadening partly and may operate at higher launch powers, which may be beneficial for higher modulation formats or reach. Even by the application of  $\gamma$ -lifting to this former AE, its performance stays well below the one proposed here, due to the limitations of its linear architectural template. The CNN-based AE is not able to reach the numerical limit that was introduced by Fig. 2, but significantly contributes to closing this gap.

## 5. Summary and conclusion

In this work, we have identified spectral broadening introduced by Kerr-nonlinearity as one of the major performance limitation of a DBP for the simulated optical fiber channel with bandlimited TX and RX.

The nowadays well-known AE was chosen as an approach to learn to address this challenge and to find an appropriate nonlinear pulse shaping. We showed that the TX-part of the AE can be trained through the SSFM. As it turned out, a stable convergence of the trainable structures requires a carefully adjusted training method. We have therefore combined a conventional training with a novel “ $\gamma$ -lifting” strategy to achieve significant gains compared to the currently best known NLC schema, i.e., split DBP, in a regime where spectral broadening is significant.

For now, we have only considered a single channel and single polarization scenario, to obtain a better understanding of the effects of Kerr-nonlinearity. In future work, these ideas can be extended to a multi-channel system (wavelength division multiplex (WDM)), or by applying even more sophisticated architectural templates.

## References

- [1] T. Uhlemann, S. Cammerer, A. Span, S. Dörner, and S. ten Brink, “Deep-learning Autoencoder for Coherent and Nonlinear Optical Communication,” in *Photonic Networks; 21st ITG-Symposium*, 2020.
- [2] D. S. Millar, S. Makovejs, C. Behrens, S. Hellerbrand, R. I. Killey, P. Bayvel, and S. J. Savory, “Mitigation of Fiber Nonlinearity Using a Digital Coherent Receiver,” *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 16, no. 5, pp. 1217–1226, Sep. 2010.
- [3] M. I. Yousefi and F. R. Kschischang, “On the Per-Sample Capacity of Nondispersive Optical Fibers,” *IEEE Transactions on Information Theory*, vol. 57, no. 11, pp. 7522–7541, 2011.
- [4] T. Xu, B. Karanov, N. A. Shevchenko, D. Lavery, G. Liga, R. I. Killey, and P. Bayvel, “Digital nonlinearity compensation in high-capacity optical communication systems considering signal spectral broadening effect,” *Scientific Reports*, vol. 7, no. 1, p. 12986, Dec. 2017.
- [5] R.-J. Essiambre, G. Raybon, and B. Mikkelsen, “Chapter 6 - Pseudo-Linear Transmission of High-Speed TDM Signals: 40 and 160 Gb/s,” in *Optical Fiber Telecommunications IV-B (Fourth Edition)*, fourth edition ed., ser. Optics and Photonics, I. P. Kaminow and T. Li, Eds. Burlington: Academic Press, 2002, pp. 232–304.
- [6] D. M. Pepper and A. Yariv, “Compensation for phase distortions in nonlinear media by phase conjugation,” *Optics Letters*, vol. 5, no. 2, p. 59, Feb. 1980.
- [7] E. Bidaki and S. Kumar, “A Raman-Pumped Dispersion and Nonlinearity Compensating Fiber For Fiber Optic Communications,” *IEEE Photonics Journal*, vol. 12, no. 1, pp. 1–17, Feb. 2020.
- [8] K. Peddanarappagari and M. Brandt-Pearce, “Volterra series transfer function of single-mode fibers,” *Journal of Lightwave Technology*, vol. 15, no. 12, pp. 2232–2241, Dec. 1997.
- [9] X. Li, X. Chen, G. Goldfarb, E. Mateo, I. Kim, F. Yaman, and G. Li, “Electronic post-compensation of WDM transmission impairments using coherent detection and digital signal processing,” *Optics Express*, vol. 16, no. 2, p. 880, 2008.
- [10] A. Mecozzi and R.-J. Essiambre, “Nonlinear Shannon Limit in Pseudolinear Coherent Systems,” *Journal of Lightwave Technology*, vol. 30, no. 12, pp. 2011–2024, Jun. 2012.
- [11] M. I. Yousefi and F. R. Kschischang, “Information Transmission Using the Nonlinear Fourier Transform, Part III: Spectrum Modulation,” *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 4346–4369, Jul. 2014.
- [12] B. Karanov, M. Chagnon, F. Thouin, T. A. Eriksson, H. Bulow, D. Lavery, P. Bayvel, and L. Schmalen, “End-to-End Deep Learning of Optical Fiber Communications,” *Journal of Lightwave Technology*, vol. 36, no. 20, pp. 4843–4855, Oct. 2018.
- [13] D. Lavery, D. Ives, G. Liga, A. Alvarado, S. J. Savory, and P. Bayvel, “The Benefit of Split Nonlinearity Compensation for Single-Channel Optical Fiber Communications,” *IEEE Photonics Technology Letters*, vol. 28, no. 17, pp. 1803–1806, Sep. 2016.
- [14] C. Häger and H. D. Pfister, “Physics-Based Deep Learning for Fiber-Optic Communication Systems,” *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 1, pp. 280–294, 2021.
- [15] V. Neskorniuk, A. Carnio, V. Bajaj, D. Marsella, S. K. Turitsyn, J. E. Prilepsky, and V. Aref, “End-to-End Deep Learning of Long-Haul Coherent Optical Fiber Communications via Regular Perturbation Model,” in *2021 European Conference on Optical Communication (ECOC)*, 2021, pp. 1–4.
- [16] A. Span, V. Aref, H. Bulow, and S. ten Brink, “Time-Bandwidth Product Perspective for Nonlinear Fourier Transform-Based Multi-Eigenvalue Soliton Transmission,” *IEEE Transactions on Communications*, vol. 67, no. 8, pp. 5544–5557, Aug. 2019.
- [17] S. Gaiarin, F. Da Ros, R. T. Jones, and D. Zibar, “End-to-End Optimization of Coherent Optical Communications Over the Split-Step Fourier Method Guided by the Nonlinear Fourier Transform Theory,” *Journal of Lightwave Technology*, vol. 39, no. 2, pp. 418–428, Jan. 2021.
- [18] G. P. Agrawal, *Nonlinear fiber optics*. Springer, 2000.
- [19] G. Kramer, M. I. Yousefi, and F. R. Kschischang, “Upper bound on the capacity of a cascade of nonlinear and noisy channels,” in *2015 IEEE Information Theory Workshop (ITW)*. Jerusalem, Israel: IEEE, Apr. 2015, pp. 1–4.
- [20] R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, “Capacity Limits of Optical Fiber Networks,” *Journal of Lightwave Technology*, vol. 28, no. 4, pp. 662–701, Feb. 2010.
- [21] K. Roberts, Chuandong Li, L. Strawczynski, M. O’Sullivan, and I. Hardcastle, “Electronic precompensation of optical nonlinearity,” *IEEE Photonics Technology Letters*, vol. 18, no. 2, pp. 403–405, Jan. 2006.
- [22] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” *arXiv:1412.6980 [cs]*, Jan. 2017, arXiv: 1412.6980.
- [23] D. Tandler, S. Dörner, S. Cammerer, and S. ten Brink, “On Recurrent Neural Networks for Sequence-based Processing in Communications,” in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*. Pacific Grove, CA, USA: IEEE, Nov. 2019, pp. 537–543.

# Time Adaptive Probabilistic Shaping for Combined Optical/THz Links

In-Ho Baek<sup>1</sup>, Frederik Bart<sup>2</sup>, Robert Elschner<sup>1</sup>, Fred Meier<sup>2</sup>,

David Hellmann<sup>2</sup>, Andreas Maaßen<sup>1</sup>, Colja Schubert<sup>1</sup>, and Ronald Freund<sup>1</sup>

<sup>1</sup>Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institute, Einsteinufer 37, 10587 Berlin, Germany

<sup>2</sup>Chair of Climatology, Institute of Ecology, Technische Universität Berlin, Rothenburgstraße 12, 12165 Berlin, Germany

Email: in-ho.baek@hhi.fraunhofer.de

**Abstract**— By exploiting the large bandwidth in the frequency band around 300 GHz, THz communications are a highly anticipated technology for future converged fiber-optical & THz-wireless access and backhaul networks because of the high achievable data rates. However, unlike in fiber-optical communications, the transmission quality over the THz wireless link is highly influenced by weather-dependent atmospheric losses, resulting in dynamically varying signal-to-noise ratios (SNR). Probabilistic amplitude shaping (PAS) has proven itself a powerful method that can adapt the signal transmission to changing channel dynamics. Additionally, signal shaping results in a gain over uniform QAM signaling, which can asymptotically amount up to 1.53 dB for arbitrarily high constellation cardinalities. We investigate the applicability of PAS for an outdoor THz wireless link in simulations with realistic models of the weather-dependent loss. The link performance is evaluated in terms of achievable information rate (AIR) and bit error ratio (BER). Optimal shaping entropies are determined to adjust error rates to a given forward error correction (FEC) threshold under a varying SNR, resulting in a tradeoff between the net data rate and the error rates. Furthermore, the optimal choice for the base constellation with respect to complexity, capacity, and reliability is evaluated. Finally, link performance statistics are evaluated using real weather data. © 2022 The Author(s)

**Keywords** — *Fiber optics, Terahertz communications, Adaptive optics*

## I. INTRODUCTION

In order to react to the increasing demand for high data rates in wireless communication networks, interest for research in the field of terahertz (THz) transmission covering the frequency band between 100 GHz and 10 THz has rapidly grown in the scientific community [1]. The ability of THz links to be seamlessly integrated into existing fiber-optic infrastructure is one of their many advantages, as it combines the flexibility and adaptability of wireless links with the reliability and capacity of fiber-optic links [2]. By sharing the infrastructure, the costs for network upgrades can be significantly reduced [3]. Consider the concept of such a combined optical/THz link depicted in Fig. 1a). By using an analog optical/RF interface on the transmit side, optical signals are converted into THz signals that are then transmitted over a line of sight wireless channel. In analogy, an RF/optical interface converts the THz signal back into the

optical domain on the receive side. However, unlike fiber-optical transmission, wireless transmission at frequencies above 100 GHz suffers from attenuations caused by atmospheric water vapor [4]. This happens as water drops resulting from rain and fog attenuate the THz waves by absorbing and scattering the incident wave. These atmospheric losses will add to the free-space path loss (FSPL) between the two antennas, resulting in weather-dependent, dynamic variations of the channel signal-to-noise ratio (SNR). Therefore, rate adaptation techniques have to be applied in order to overcome the challenges caused by the dynamic SNR.

In recent years, probabilistic amplitude shaping (PAS) [5] has been extensively studied not only for the sake of transmission distance extension or increasing the spectral efficiency at a fixed SNR, but also for rate adaptivity [6]. Depending on the shaping entropy, i.e. the number of bits per transmit symbol, PAS changes the symbol distribution by increasing the probability of occurrence of low-power symbols, corresponding to inner constellation points, and reducing the probability of occurrence of high-power symbols, corresponding to outer constellation points. As it is ideally suited for the additive white Gaussian noise (AWGN) channel, PAS modulation can achieve higher information rates at the same SNR than uniform quadrature-amplitude modulation (QAM) with the same entropy.

However, while the achievable information rates only correspond to an upper limit assuming an ideal, capacity-achieving forward-error correction (FEC) with flexible overhead, a real FEC implementation is typically constrained to a fixed overhead, which results in a fixed pre-FEC bit error ratio (BER) threshold. To feasibly integrate rate-adaptive PAS with a typical soft-decision (SD)-FEC code, its pre-FEC BER threshold must not be exceeded during adaptation. This is not the case when choosing the shaping entropy in order to maximize the achievable information rate. It is therefore of interest to study the information rates of PAS under a given pre-FEC BER threshold.

This paper reports on the simulation and the evaluation of a rate-adaptive THz-wireless communication link using PAS, where the wireless THz channel is modeled by an AWGN channel with dynamically changing SNR derived from real weather data. We use a simple lookup-table-based method to keep the pre-FEC BER below a given threshold during adaptation, and analyze achievable information rates and reliabilities of the THz-wireless link. The results obtained with PAS are compared to those acquired with constellation switching between uniform QAM formats.

This work has been funded by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF) under the AI-NET-PROTECT (16KIS1282) project and project 6G SENTINEL funded by the Fraunhofer Gesellschaft e.V. für Angewandte Forschung.

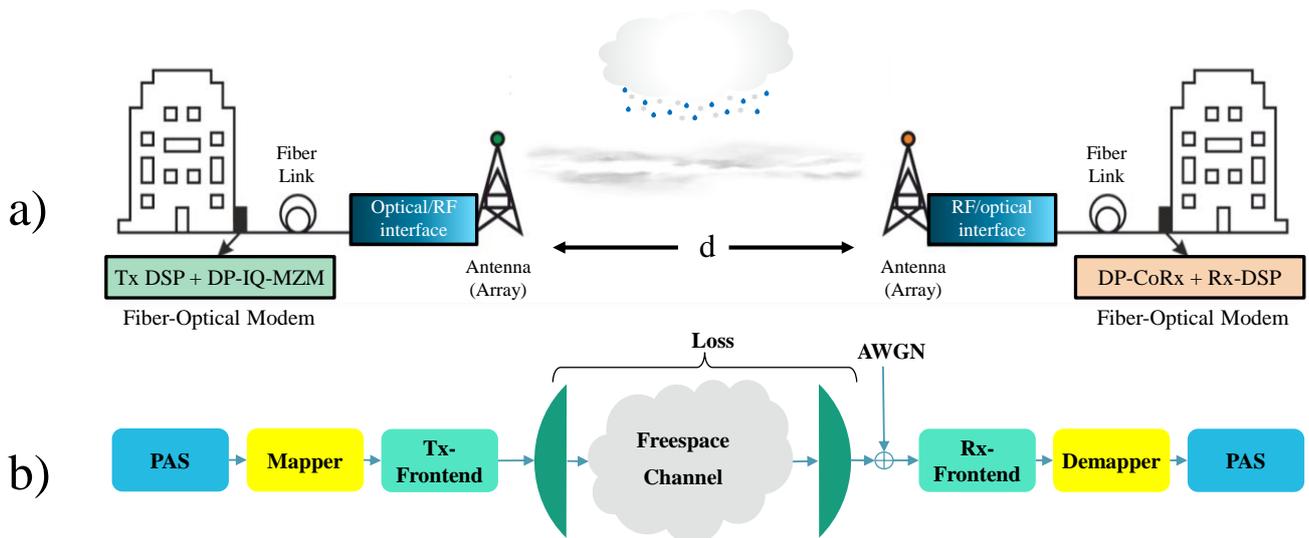


Fig. 1 a) Concept of a unidirectional THz-wireless fiber extender. The optical/RF interfaces comprise analog coherent baseband frontends that allow the direct conversion from the fiber-optical domain to the electrical base-ban and further to the THz-wireless domain without local DSP, and vice versa. b) Simplified blockdiagram of the simulated environment. PAS: Probabilistic Amplitude Shaping.

This paper is organized in the following way: In Sec. II, an overview of the simulation setup and the parameters assumed for the link are introduced. Sec. III deals with the evaluation of the link capacity obtained through real weather data, as well as the calculation of the link SNR. In Sec. IV, we briefly introduce the concepts of PAS and constellation switching. We also cover the adjustment of the shaping entropy to fixed forward error correction (FEC) thresholds. In Sec. V, we evaluate the results and discuss the performance of the different rate-adaptation schemes using pre-FEC BER, AIR, and cumulative capacity as the figure of metrics. This paper closes with a summary and conclusion in Sec. VI.

## II. SETUP AND LINK PARAMETERS

Fig. 1b) shows the structure of the simulated THz system. On the transmitter side, the PAS block maps independent and identically distributed random bits onto probabilistically shaped amplitudes and back into the bit level via a constant composition distribution matcher (CCDM) (see Sec. IV). A mapper subsequently maps these bits onto symbols taken from the alphabet of the considered base constellation. For each simulation run, we generated  $2 \cdot 10^4$  probabilistically shaped symbols. Via the Tx frontend, these symbols are transmitted over an AWGN channel with a given SNR. By repeating the simulation for each time event in a set of real weather data, we account for the SNR dynamics caused by weather-dependent losses as well as for the static free-space path losses (FSPL). We determine the weather effects based on attenuation models by the International Telecommunication Union (ITU). The Rx frontend receives the noisy signal before the demapper block maps the symbols back to bits. The demapping process takes into account the

Tab. 1 Parameters of the THz link.

Carrier Frequency $f$	270 GHz
Symbol rate $R_{sym}$	32 GBd
Transmit Power $P_{Tx}$	0 dBm
Tx Antenna Gain $G_{Tx}$	55 dBi
Noise Figure $NF$	10 dB
Rx Antenna Gain $G_{Rx}$	55 dBi
Distance $d$	1000 m

a-priori probability of the transmitted PAS symbols, resulting in a so-called maximum a posteriori (MAP) detection [7]. Finally, the generalized mutual information (GMI) and pre-FEC BER is used to assess the system performance. In this case, the GMI is calculated based on the input distribution and the channel transition probabilities, which are obtained using bit-level log-likelihood ratios (LLR). For bit-interleaved coded modulation (BICM), which is the coding scheme used for our work, the GMI is a suitable metric of AIR [8]. Both the GMI and AIR are ultimately limited by the Shannon capacity of the channel.

Tab. 1 lists the link parameters which were adopted from recent experiments in [2] and used in the simulations. The distance between the two THz antennas is assumed to be 1000 m, until stated otherwise, while the linear transmit power is 0 dBm at a carrier frequency of 270 GHz and a symbol rate of 32 GBd. Further, highly directive Tx and Rx antennae with gains of 55 dBi each are assumed on both link ends. The receiver frontend is assumed to have a noise figure (NF) of 10 dB. For these parameters, the resulting channel SNR only considering FSPL is 27.8 dB.

## III. CAPACITY ANALYSIS

As discussed in the previous section, the SNR of the simulated AWGN channels is calculated using actual weather data. In this section, we will give a brief overview of the used data as well as the calculation of the weather-dependent losses. In addition, we show how to calculate the overall link SNR from these losses.

### A. Weather data

In order to evaluate the link capacity under realistic conditions, the weather-dependent attenuation was calculated using stationary measurement data by the German Meteorological Service (Deutscher Wetterdienst, DWD). This 10-year dataset includes the parameters air temperature, relative humidity, air pressure, and precipitation rate from 2011 until 2020. Parameters which were only available in a ten-minute resolution were linearly interpolated to allow for a temporal resolution of 1 minute. This was necessary to

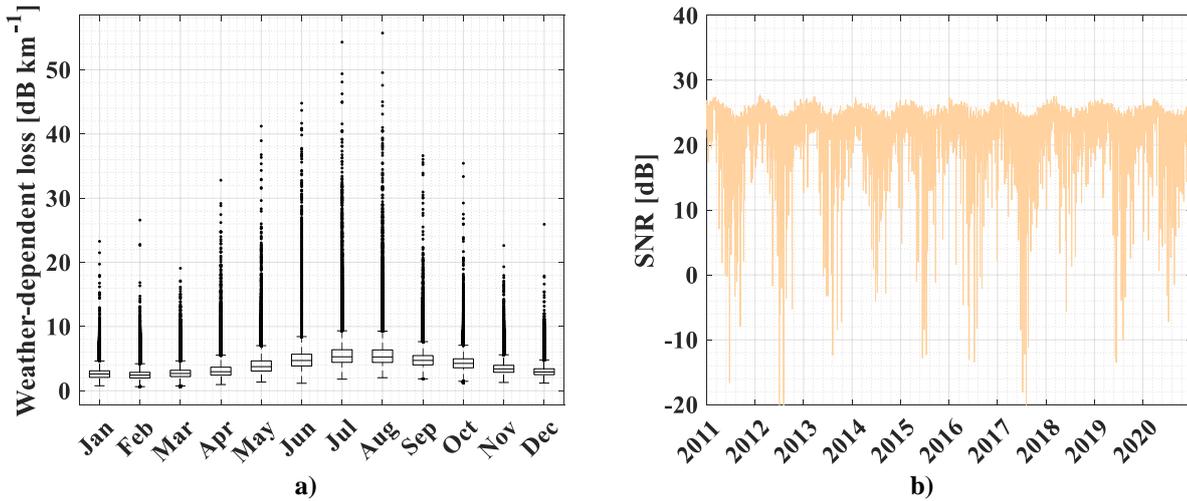


Fig. 2 a) Data distribution for the calculated weather dependent loss due to atm. gases and rain categorized by month. The corresponding weather data were measured at the DWD weather station at Berlin Tempelhof from 2011 to 2020. The data is visualized as a boxplot, categorized by months. The interquartile range (IQR) or the middle 50% of the data is represented as a box with the horizontal line in the middle, indicating the median value. Observations beyond the whisker lengths of 1.5 IQR are marked as outliers. The variability is highest during the summer. The highest outliers were observed during July and June resulting in attenuation values of up to  $\sim 55$  dB km<sup>-1</sup>. b) SNR calculated over a ten-year period for Berlin Tempelhof for an antenna spacing of  $d=1000$  m. A yearly periodicity is observed, with SNR values dropping below -10 dB occasionally during the summer.

allow for a better representation of short-term changes in precipitation. Any gaps in the time series due to missing values were also approximated using linear interpolation. The calculation of the absolute atmospheric humidity  $p_v$  (kg m<sup>-3</sup>) was based on the ideal gas law for pure water vapor:

$$p_v = \frac{e}{R_v T}, \quad (1)$$

with  $T$  for the air temperature (K),  $R_v$  for the specific gas constant for water vapor (461.5 J kg<sup>-1</sup> K<sup>-1</sup>), and  $e$  for the partial water vapor pressure (Pa) which was determined as follows:

$$e = E RH, \quad (2)$$

where  $RH$  is the relative humidity as a dimensionless value and  $E$  is the saturation vapor pressure (hPa) which was calculated using the empirical equations (3) and (4) based on the approximation by H.G. Magnus [9],

$$E(T) = 6.1078 \text{ hPa} \exp\left(\frac{17.08085 T [\text{°C}]}{234.175 \text{ °C} + T [\text{°C}]}\right) \quad (3)$$

Tab. 2 Quantiles of weather-dependent loss (dB km<sup>-1</sup>) due to atm. gases and rain for 10 years at Berlin Tempelhof (2011-2020). The year 2017 showed a noticeably higher attenuation for the 99.99% quantile than other years.

Percentage (%)	99	99.9	99.99
2011	8.5	12.5	21.9
2012	8.5	12.2	25.5
2013	8.6	12.1	21.3
2014	8.4	11.3	18.4
2015	8.2	11.7	20.0
2016	7.8	11.3	24.1
2017	8.6	14.7	25.8
2018	7.5	10.3	18.0
2019	8.1	11.6	25.4
2020	8.2	11.8	19.2
Total	8.3	12.1	22.1

assuming a water surface for negative  $T$  [°C], and

$$E(T) = 6.1078 \text{ hPa} \exp\left(\frac{17.84362 T [\text{°C}]}{245.425 \text{ °C} + T [\text{°C}]}\right) \quad (4)$$

assuming an ice surface for positive  $T$  [°C].

#### B. Weather dependent loss

The weather-dependent loss was calculated as the sum of the specific gaseous attenuation and the rain-induced attenuation. The propagation of electromagnetic waves from THz transmission is strongly affected by the atmospheric composition. Gas molecules (especially oxygen and water vapor) can inhibit the oscillation of these waves and reduce the signal strength. The total gaseous attenuation  $A_{gas}$  (dB km<sup>-1</sup>) was calculated based on the ITU-R recommendation P.676.12 [10]:

$$A_{gas} = \gamma_G r_0 = (\gamma_o + \gamma_w) r_0, \quad (5)$$

where  $\gamma_G$  is the specific gaseous attenuation,  $\gamma_o$  is the specific attenuation due to dry air (oxygen, pressure-induced nitrogen, and non-resonant Debye attenuation),  $\gamma_w$  is the specific attenuation due to water vapor and  $r_0$  is the path length (km).

As another important factor, precipitation can significantly impact the THz-wireless link performance due to Mie scattering effects along the transmission path. The total rain-induced attenuation  $A_{rain}$  (dB km<sup>-1</sup>) of the THz signal was calculated based on the ITU-R recommendation P.838-3 [11].

$$A_{rain} = \gamma_R r_0 = k R^\alpha r_0, \quad (6)$$

where  $\gamma_R$  is the specific rain-induced attenuation,  $R$  is the precipitation rate (mm h<sup>-1</sup>) and  $\alpha$  and  $k$  are frequency dependent coefficients for the linear and circular polarization which can be calculated using the tables in ITU-R recommendation P.838-3 [11].

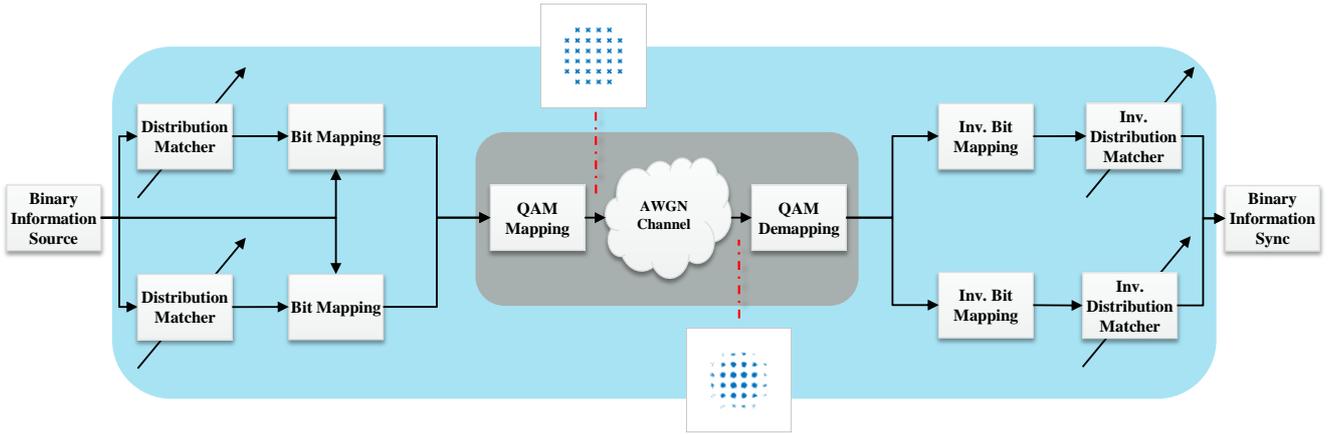


Fig. 3 Schematic description of probabilistic amplitude shaping. The elements contained in the gray box are basic components of any QAM based transmission over an AWGN channel, while the blue box contains the further needed elements to enable PAS assisted transmission.

The attenuation due to clouds and fog was not taken into account in this study since accurate measurement data of the liquid water density for clouds or fog was not available. Additionally, the corresponding ITU-R recommendation P.840-8 [12] states, that the Rayleigh approximation for cloud and fog is only valid for frequencies up to 200 GHz. In fact, it was shown in [13] that the Rayleigh approximation considerably overestimates the fog attenuation at frequencies  $> 200$  GHz in comparison to experimental observations and that the propagation of THz pulses through dense fog performs similarly as for the reference case without fog.

The data distribution of the weather-dependent loss based on the 10-year DWD data is visualized as a boxplot, categorized by months in Fig. 2a). The interquartile range (IQR) or the middle 50% of the data is represented as a box with the horizontal line in the middle, indicating the median value. For most months (especially during the early year and winter) the IQR is relatively narrow between 2 to 4 dB km<sup>-1</sup>. However, during the summer months (from June until August), the data variability was typically higher with a  $\sim 60$  % larger IQR on average, compared to all other months. For these three months the median attenuation was on average  $\sim 50$  % higher than during the rest of the year. In addition, the highest outliers (shown as black dots) were observed during this season in all years with attenuations up to  $\sim 55$  dB km<sup>-1</sup>. This could be explained due to the higher average humidity and more frequent heavy precipitation events during the summer months.

To further investigate the link availability, quantiles above 99 % were calculated for individual years as well as the complete 10-year dataset in Tab. 2. These values represent the modeled attenuation below of which 99 %, 99.9 % and 99.99 % of the data falls. Except for the 99.99 % quantile almost all years show similar results with only slight differences. However, for 2017 the attenuation for the 99.9 % quantile was noticeably higher than the other years. Here, the attenuation for the 99.99% quantile was also the highest, although the difference to the other years was much smaller. This was mainly attributed to the unusually heavy and prolonged precipitation events during this year. These results suggest that link availability was the lowest during 2017.

### C. SNR calculation

The obtained attenuations can further be processed to yield an overall SNR of the THz wireless link. First, we estimate the received power on the receive side using the Friis transmission formula

$$P_{rx} = P_{tx} - L_A + G_{Tx} + G_{Rx}, \quad (7)$$

where  $L_A$  is the resulting loss from atmospheric effects and the free space path loss

$$L_A = (A_{gas} + A_{rain}) * \frac{d}{1000} + FSPL, \quad (8)$$

with  $FSPL = \left(\frac{4\pi df}{c}\right)^2$ , where  $c$  is the speed of light in a vacuum. The overall channel SNR is then estimated by

$$SNR_{dB} = P_{rx} - P_{noise}, \quad (9)$$

with  $P_{noise} = kTB_{dB} + NF$ , where  $kTB_{dB}$  is the thermal noise power in dB and NF the noise figure.

The obtained SNR values for the 10-year dataset are displayed in Fig. 2b). It is evident that there is a yearly periodicity in the SNR dynamics, with the highest degradation usually occurring during the summer.

## IV. OPERATING PARAMETER OF PAS

In this section, we first introduce the probabilistic amplitude scheme. Then, the optimal shaping entropies w.r.t. the AIR and BER performance are presented. Finally, a reference method for rate adaptivity, namely constellation switching, is introduced for comparison.

### A. Introduction to PAS

In a nutshell, the idea of probabilistic shaping is the transmission of regular QAM constellations with different occurrence probabilities of the symbols. As the optimal constellation for an AWGN channel is a Gaussian, equiprobable uniform QAM constellations are suboptimal and have an asymptotic SNR gap of  $\pi e/6 \approx 1.53$  dB as compared with capacity-achieving Gaussian constellations, where  $e$  is Euler's number. [15]. In the case of an AWGN channel, the so-called Maxwell-Boltzmann (M-B) distribution has been shown to provide the optimal probability distribution for regular QAM constellations [16],

which allows a performance close to the Shannon limit. The probability of a constellation point  $x_i$  being transmitted is then

$$P(x_i) \propto \exp(-\lambda|x_i|^2), \quad (10)$$

where  $\lambda$  represents the amount of the applied shaping [3], i.e.  $\lambda = 0$  would result in the uniform signal. The matching of random input bits onto probabilistically shaped output constellations is accomplished by using a distribution matcher (DM). In its essence, the DM is used to convert  $k$  uniformly distributed bits  $d^k = d_1 d_2 \dots d_k$  with  $d \in \{0,1\}$  from a binary information source into a non-uniformly distributed sequence of  $n$  amplitudes  $a^n = a_1 a_2 \dots a_n$ . This process is invertible, i.e. amplitudes  $a^n$  mapped back onto  $d^k$  recovers the original bitstream. The amount of bits contained in each of these amplitudes is the so-called DM rate

$$R = \frac{k}{n} \left[ \frac{\text{Bits}}{\text{Output symbol}} \right] \quad (11)$$

and sets the ground for the desired output target distribution. For large  $n$ , this rate approaches the desired shaping entropy

$$H(X) = -\sum_i P(x_i) \log_2 P(x_i), \quad (12)$$

where  $X$  is the constellation alphabet. These amplitudes are then Gray-mapped back into the bit-level, resulting in a non-uniform unipolar PAM constellation. More specifically, each amplitude is assigned to  $\left(\frac{\log_2 M}{2} - 1\right)$  bits with binary reflected Gray code (BRGC) labeling, where  $M$  is the cardinality of our target rectangular QAM constellation. A bipolar pulse-amplitude-modulated (PAM) constellation is obtained by adding equiprobable sign bits, which represent signs that are uniform on  $\{-1, +1\}$ , to each of the unipolar PAM symbols. Due to the uniform distribution of the sign bits, the resulting constellation still follows the desired M-B distribution. Several distribution matchers have been proposed in the literature, which can generally be divided into two groups, namely variable-length matchers and fixed-length DMs. For our simulations, we consider a fixed-length DM, based on the constant composition distribution matching (CCDM) [17] scheme. Due to the fact that any rectangular M-QAM constellation can be expressed as a Cartesian multiplication of an in-phase arm and a quadrature arm PAM constellation, combining two orthogonal bipolar PS-PAMs results in an equivalent rectangular PS-QAM constellation, provided that DMs with the same properties for each PAM constellation are used. The general scheme of PAS used in our simulations is demonstrated in Fig. 3. The two arms after the binary information source illustrate the DM process for each in-phase and quadrature PAM, whereas the QAM mapping block combines these two orthogonal PAM constellations. After the transmission over the AWGN channel with a given SNR, the received QAM symbols are mapped back into bits. Subsequently, performance evaluation metrics, such as the pre-FEC BER and AIR, are assessed. The inverse CCDM is shown only for illustrative purpose, but was not used in the evaluations.

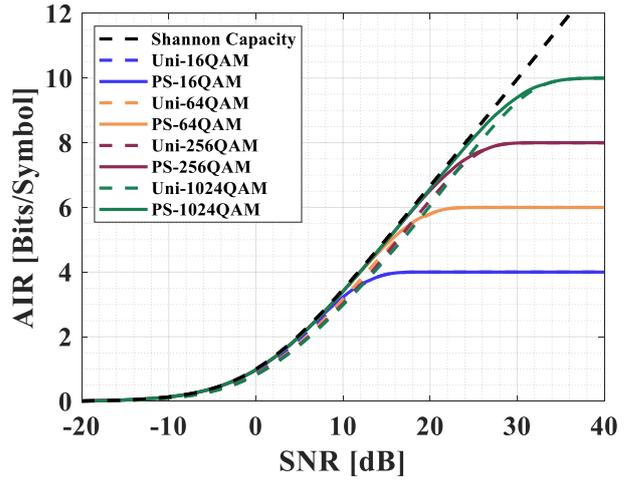


Fig. 4 Achievable information rates as functions of the SNR for uniformly-distributed constellations (dashed lines) and probabilistically-shaped constellations (solid lines) with optimal  $\lambda$  over the AWGN channel.

In order to simulatively obtain the optimum shaping entropies for a given SNR, a 2D sweep over the SNR (with a 0.1 dB step size) and over all shaping entropies the range of  $[2, \log_2 M]$  bits/symbol (with a step size of 0.1 bits) was done. By this, a lookup table (LUT) containing all the AIRs and pre-FEC BERs for each sweep is obtained for each considered modulation format. In Fig. 4, we present the AIRs for uniform formats and their probabilistically shaped counterparts. The graphs of the PS-MQAM formats show the resulting AIRs when choosing  $\lambda$  in a way that the AIR gets maximized, i.e. rate-maximizing distributions are applied by the DM for each SNR. This means that there exists an optimal value  $\lambda_{\text{opt}}$  of the shaping entropy for each SNR that maximizes the overall AIR [18]. The PS- and uniform QAM modulation formats perform similarly at high SNR: PS-1024QAM and Uni-1024QAM show an AIR saturation at SNR values of about 35 dB, PS-256QAM and Uni-256QAM at about 28 dB, PS-64QAM and Uni-64QAM at about 22 dB and PS-16QAM and Uni-16QAM at about 15 dB. For SNRs below that saturation region, we observe a significantly better AIR of the PS-QAM formats which closely approach the Shannon capacity. For SNR values close to and below 0 dB, all modulation formats achieve a similar AIR.

### B. Optimal shaping entropies

Based on numerical simulations of PS-1024QAM, we evaluated the pre-FEC BER and AIR performance as a function of the SNR with rate-maximizing shaping, while considering the SNR range of  $[-20, 40]$  dB. The maximal AIR values are iteratively searched in the obtained LUT and indexed accordingly. Subsequently, the corresponding BERs are extrapolated from the indexed AIRs. The resulting BER values are plotted in Fig. 5a). It should be noted here that every PS-M-QAM format is limited by two corner cases:

- The lower bound is the BER performance of the signal with an entropy of 2 bit/symbol, as maximal shaping  $\lambda_{\text{max}}$  of a square M-QAM always results in a quadrature-phase-shifted signal (QPSK).

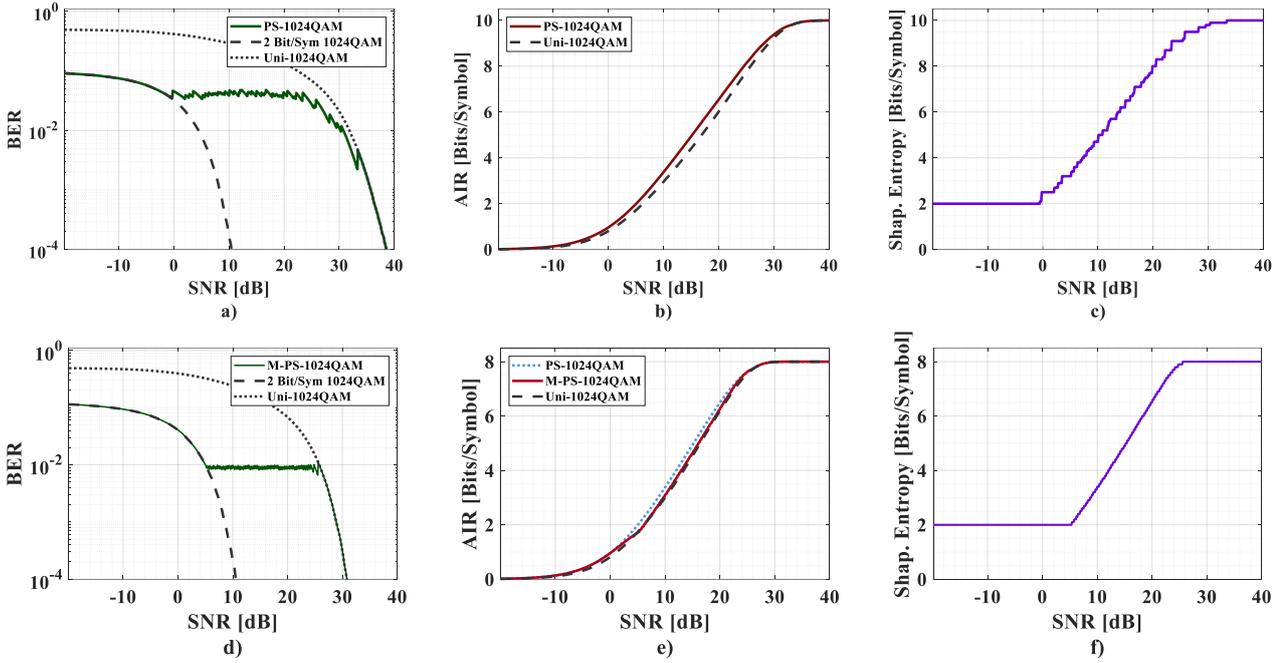


Fig. 5. a) BER performance of PS-1024QAM limited by the bordercases  $\lambda_{max}$  and  $\lambda = 0$ ; b) Achievable information rates of PS-1024QAM and Uni-1024QAM; c) Corresponding shaping entropies for PS-1024QAM d) BER performance of M-PS-1024QAM limited by the bordercases  $\lambda_{max}$  and  $\lambda = 0$ ; e) Achievable information rates of M-PS-1024QAM, PS-1024QAM and Uni-1024QAM; f) Corresponding shaping entropies for M-PS-1024QAM.

- The upper bound is given for  $\lambda = 0$ , i.e. no shaping, resulting in the BER performance of the uniform version of the same cardinality.

Between these two corner cases, i.e. at approximately 0 dB SNR, the BER of the PS-1024QAM separates from the BER curve of the case where  $\lambda = \lambda_{max}$ , and stays nearly constant until it adheres to the BER curve for the case  $\lambda = 0$  at about 32 dB SNR. We refer to the quasi-constant region as the BER plateau, which is located at around  $4 \cdot 10^{-2}$  BER. Fig. 5c) shows the corresponding shaping entropies that were used for the shaping. This indicates that, in order to keep the AIR in the quasi-linear region closest to the Shannon limit, the BER is held almost constant at  $4 \cdot 10^{-2}$  for the case of PS-1024QAM. While this method results in overall transmission with the highest AIR, a pre-FEC BER at about  $4 \cdot 10^{-2}$  is too high for a typical implementation of SD-FEC schemes at a reasonable overhead. It is therefore of interest to lower the BER plateau until it matches a given FEC threshold, thus trading a better BER performance for a small rate loss.

To make sure that the BER plateau matches the desired threshold, the indexing order in the LUT is reversed as compared to the optimal AIR indexing, i.e. the BER values

matching the threshold are indexed and the corresponding AIRs are extrapolated. In Fig. 5d), results for a BER threshold of  $1 \cdot 10^{-2}$  are shown for the format designated as M-PS-1024QAM. Through the aforementioned process of reverse indexing, a much more stable plateau is observed. As expected, we can notice a small rate loss in contrast to the rate maximizing shaping, which amounts up to approximately 0.6 dB in the linear regime, as can be seen in Fig. 5e). It is also clear that in order to achieve lower BER values, stronger shaping must be applied at all SNR, shifting the shaping entropy curve to the right, as depicted in Fig. 5f).

### C. Constellation switching

For comparison, we consider an alternative method that adapts the rate to the channel SNR by using a technique called constellation switching (CS). In this technique, rate adaptivity is achieved by changing the modulation format cardinality at the transmitter depending on the channel SNR. In our simulations, CS-1024QAM can switch between square 4-, 16-, 64-, 256-, and 1024-QAM. The possible analogous signal entropies are therefore 2, 4, 6, 8, and 10 bits per symbol. In analogy to M-PS-MQAM, a LUT is created for the uniform format of each cardinality. To match the BER

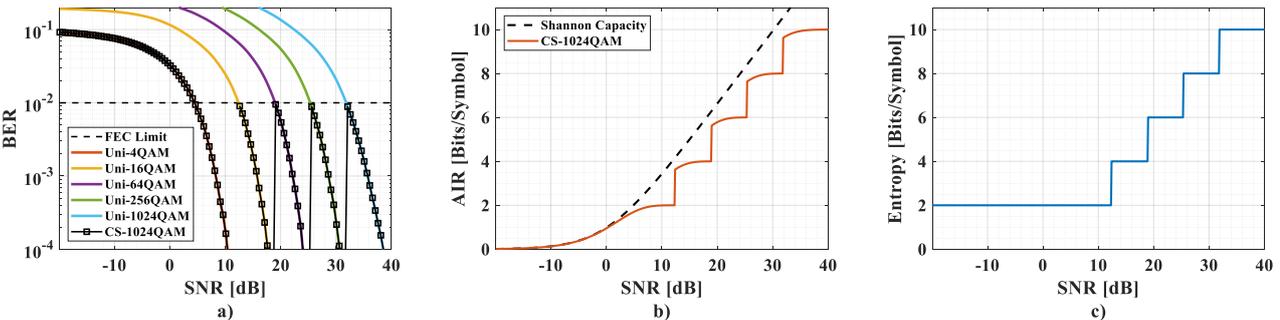


Fig. 6 a) BER performance of CS-1024QAM and of uniform formats; b) Achievable information rates for CS-1024QAM w.r.t. the Shannon capacity; c) Entropy of CS-1024QAM. The CS-1024QAM performance is shown in all results using a switching BER threshold of  $1 \cdot 10^{-2}$ .

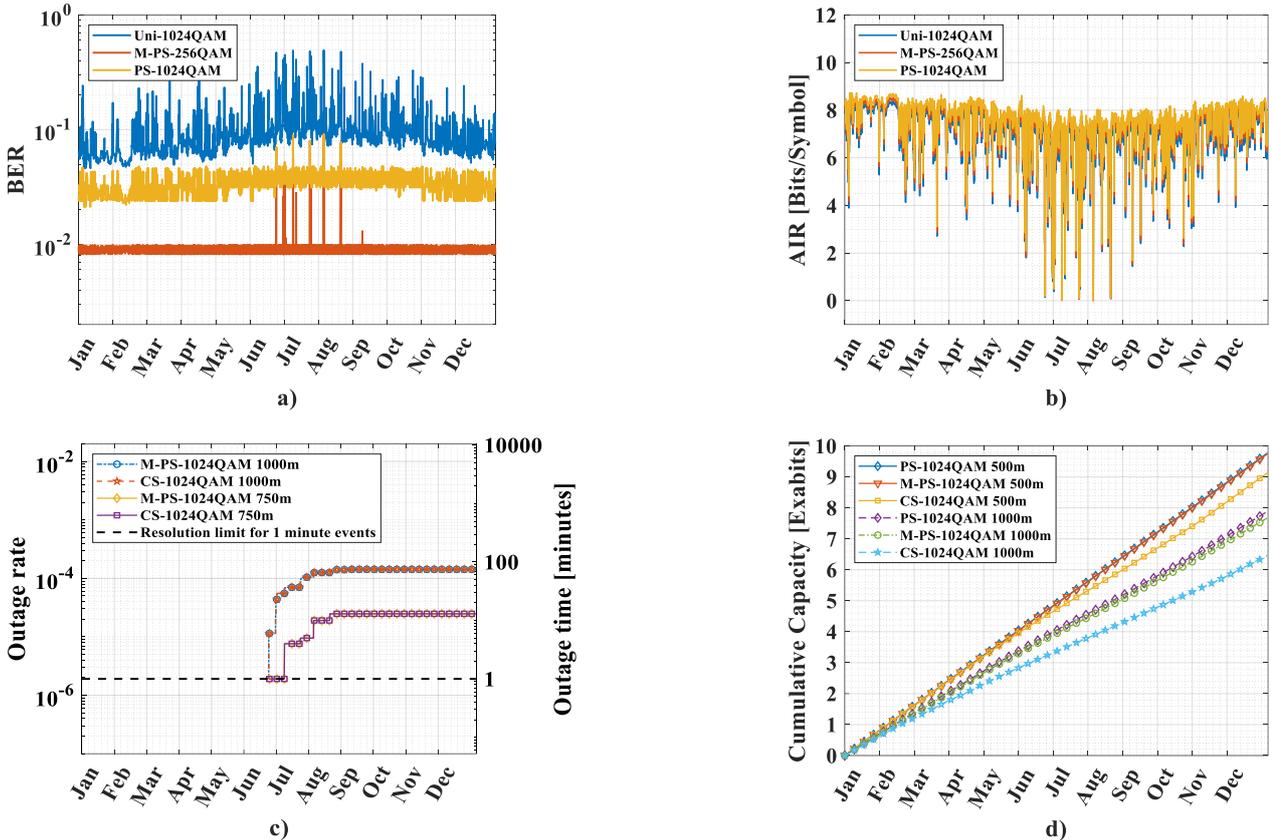


Fig. 7. a) BER performance of Uni-, PS- and M-PS-1024QAM; b) AIR performance of Uni-, PS- and M-PS-1024QAM; c) Outage rates of M-PS- and CS-1024QAM for  $d=1000\text{m}$  and  $d=750\text{m}$ ; d) Cumulative capacities of PS-, M-PS- and CS-1024QAM for  $d=1000\text{m}$ ,  $d=750\text{m}$  and  $d=500\text{m}$ . All results are shown for the exemplary year 2017.

performance to the threshold, corresponding BER values are iteratively searched and indexed. The switching from one modulation cardinality to the following one happens at the instance where the BER of the next highest modulation format falls below the set FEC threshold.

Based on the simulation environment introduced in Sec II, we evaluated the rate-adaptive transmission with CS-1024QAM. The BER and AIR performance as a function of the SNR are shown in Fig. 6a) and Fig. 6b), respectively, with a considered switching BER threshold of  $1 \cdot 10^{-2}$ .

The switching between the modulation formats happens at SNR values of approximately 12.4 dB, 19 dB, 25.3 dB, and 31.8 dB for 4-, 16-, 64-, 256- and 1024QAM respectively. Further, we observe significant rate gaps compared to the Shannon limit in the region between the modulation formats of different cardinalities.

## V. EVALUATION

This section compares the performance of PS-, M-PS-, and CS-QAM modulation formats to a uniform QAM constellation using real weather data. The evaluation is based on the metrics pre-FEC BER, AIR, and cumulative capacity over an exemplary year. Our 10-year data set showed a particularly fluctuant behavior for the year 2017, which is why we chose that year to illustrate our results.

As shown in Fig. 7a), the M-PS-1024QAM and PS-1024QAM constellations outperform Uni-1024QAM in terms of BER. Furthermore, even for M-PS-1024QAM, a small number of outages can be observed during the summer and late summer months. We define an outage as a case

where the pre-FEC BER exceeds  $1 \cdot 10^{-2}$ , which corresponds to cases where the AIR drops below 2 bits per symbol. As can be seen from both Fig. 5d) and Fig. 6a) this is the case for an  $\text{SNR} \leq 5$  dB, where even shaping to QPSK is not sufficient anymore. As discussed in Sec. II, for our simulations parameters from Tab. 1, the maximum SNR is 27.8 dB, considering only the FSPL. Thus, an outage results for a weather-dependent loss larger than 22.8 dB. Tab. 2 shows that the expected outage rate is in the order of  $1 \cdot 10^{-4}$ . This corresponds well to the result  $1.4 \cdot 10^{-4}$  for the outage rate of the 1000 m link in Fig. 7c). For 750 m the outage rate is  $2.5 \cdot 10^{-5}$ . The corresponding outage times are about 75 minutes and 13 minutes per year for 1000 m and 750 m, respectively. M-PS-1024QAM and CS-1024QAM share the Tab. 3 Average yearly capacity loss of different modulation formats as compared to the average cumulated channel capacity per year (Shannon capacity).

Modulation format	Average yearly capacity loss compared to the Shannon capacity for different distances		
	1000m	750m	500m
PS-1024QAM (*)	1.7%	3.3%	8.8%
M-PS-1024QAM (†)	4.3%	4.6%	9.0%
CS-1024QAM (†)	14.7%	14.7%	19.4%
PS-256QAM (*)	6.7%	13.6%	24.7%
M-PS-256QAM (†)	7.4%	13.7%	24.7%
PS-64QAM (*)	24.9%	34.2%	43.5%

(\*) assuming a hypothetical capacity achieving FEC with variable threshold  
 (†) assuming a fixed FEC threshold at  $1 \cdot 10^{-2}$

same outage probability. PS and uniform formats are not shown, as these do not perform below the desired pre-FEC BER threshold. At  $d = 500$  m, the outage probabilities for CS-1024QAM and M-PS-1024QAM fall below the resolution limit for 1-minute events.

Fig. 7d) presents the cumulative capacity over the year for three distances, 500 m, 750 m, and 1000 m, for each of the considered three modulation formats. It is interesting to note that, despite the rate advantage of PS-1024QAM over M-PS-1024QAM, the overall capacity gain after the sample year amounts to only 2.6 % for  $d = 1000$  m. On the other hand, PS-1024QAM over CS-1024QAM has a capacity gain of 18.0 %. These gains decrease with decreasing length of the link, i.e. 1.4 % at 750 m and 0.3 % at 500 m for PS-1024QAM over M-PS-1024QAM, and 11.78 % at 750 m and 6.56 % at 500 m for PS-1024QAM over CS-1024QAM.

Tab. 3 compares the performance of various modulation formats based on the average yearly capacity loss as compared to the cumulated channel capacity. It may seem counterintuitive that the performance gap increases with decreasing distances, but this is to be expected as the Shannon capacity increases for shorter distances as well. Furthermore, it should be noted that for the PS-MQAM formats a hypothetical capacity-achieving FEC with a variable BER threshold is assumed, while for CS-MQAM and M-PS-MQAM formats a fixed FEC threshold at  $1 \cdot 10^{-2}$  was assumed.

Finally, the impact of the base constellation size is taken into account. At  $d = 1000$  m, even a M-PS-256QAM provides better capacities than a CS-1024QAM, while the advantage decreases for  $d = 750$  m and eventually disappears for  $d = 500$  m. The latter effect results from the higher maximum AIRs of higher cardinality formats. With longer distances, however, this result indicates that implementation complexity caused by high modulation cardinalities might be avoided by using M-PS formats with lower cardinalities.

## VI. CONCLUSION

In this study, we simulatively investigated the statistics of the weather-dependent losses of a THz-wireless link using actual meteorological data collected over the past 10 years. Based on our analysis, the average weather-dependent loss for the 99 %, 99.9 %, 99.99 % quantile is estimated to be 8.3 dB km<sup>-1</sup>, 12.1 dB km<sup>-1</sup>, and 22.1 dB km<sup>-1</sup>, respectively. We were able to observe a clear periodic pattern, with the strongest attenuations and highest variations occurring in the summer months. As a way of adapting to these dynamically changing channel attenuations, probabilistic amplitude shaping has been investigated and compared with constellation switching CS-MQAM in terms of AIR and BER. Furthermore, we have compared the performance of AIR maximizing PS-MQAM formats assuming an ideal capacity-achieving FEC with variable overhead with the method M-PS-MQAM, which adapts PAS to a fixed FEC threshold based on a simple LUT-based strategy. This allows the operation of M-PS-MQAM with a real-world implemented SD-FEC while only suffering from a small yearly capacity loss of about 2.6% for a link distance of 1000 m as compared to PS-MQAM. As a reference, CS-MQAM suffers a capacity loss of about 18.0%. Moreover, we showed that the outage rates of M-PS-MQAM and of CS-MQAM are

the same, i.e.  $1.4 \cdot 10^{-4}$  corresponding to an outage time of about 75 minutes for 1000 m under our simulation conditions. Furthermore, we show that M-PS formats of lower cardinalities, e.g. M-PS-256QAM, only suffer from small performance loss as compared to M-PS-1024QAM for a link distance of 1000m. As these results are obtained from simulations, experimental validation will be required in future work.

## REFERENCES

- [1] B. Globisch, S. Nellen, L. Liebermeister, R. B. Kohlhaas, and M. Schell, "Photonic Terahertz Solutions for Sensing, Spectroscopy and Wireless Communication," in *Optical Fiber Communication Conference (OFC) 2019 (2019)*, paper Tu2D.2, Mar. 2019, p. Tu2D.2. doi: 10.1364/OFC.2019.Tu2D.2.
- [2] C. Castro, R. Elschner, T. Merkle, C. Schubert, and R. Freund, "Experimental Demonstrations of High-Capacity THz-Wireless Transmission Systems for Beyond 5G," *IEEE Commun. Mag.*, vol. 58, no. 11, Art. no. 11, Nov. 2020, doi: 10.1109/MCOM.001.2000306.
- [3] S. Wang *et al.*, "26.8-m THz wireless transmission of probabilistic shaping 16-QAM-OFDM signals," *APL Photonics*, vol. 5, no. 5, Art. no. 5, May 2020, doi: 10.1063/5.0003998.
- [4] E.-B. Moon, T.-I. Jeon, and D. R. Grischkowsky, "Long-Path THz-TDS Atmospheric Measurements Between Buildings," *IEEE Trans. Terahertz Sci. Technol.*, vol. 5, no. 5, pp. 742–750, Sep. 2015, doi: 10.1109/TTHZ.2015.2443491.
- [5] G. Bocherer, F. Steiner, and P. Schulte, "Bandwidth Efficient and Rate-Matched Low-Density Parity-Check Coded Modulation," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 4651–4665, Dec. 2015, doi: 10.1109/TCOMM.2015.2494016.
- [6] F. P. Guiomar *et al.*, "High-capacity and rain-resilient free-space optics link enabled by time-adaptive probabilistic shaping," in *45th European Conference on Optical Communication (ECOC 2019)*, Dublin, Ireland, 2019, p. 130 (4 pp.)-130 (4 pp.). doi: 10.1049/cp.2019.0864.
- [7] S. Hu *et al.*, "MAP Detection of Probabilistically Shaped Constellations in Optical Fiber Transmissions," in *Optical Fiber Communication Conference (OFC) 2019*, San Diego, California, 2019, p. W1D.3. doi: 10.1364/OFC.2019.W1D.3.
- [8] A. Alvarado, T. Fehenberger, B. Chen, and F. M. J. Willems, "Achievable Information Rates for Fiber Optics: Applications and Computations," *J. Light. Technol.*, vol. 36, no. 2, pp. 424–439, Jan. 2018, doi: 10.1109/JLT.2017.2786351.
- [9] G. Magnus, "Versuche über die Spannkraft des Wasserdampfes," *Ann. Phys.*, vol. 137, no. 2, pp. 225–247, 1844, doi: 10.1002/andp.18441370202.
- [10] International Telecommunication Union (ITU), Ed., "Recommendation ITU-R P.676-12 - Attenuation by atmospheric gases and related effects." 2019.
- [11] International Telecommunication Union (ITU), Ed., "Recommendation ITU-R P.838-3 - Specific attenuation model for rain for use in prediction methods." 2005.
- [12] International Telecommunication Union (ITU), Ed., "Recommendation ITU-R P.840-8 - Attenuation due to clouds and fog." 2019.
- [13] Y. Yang, M. Mandehgar, and D. R. Grischkowsky, "Broadband THz Signals Propagate Through Dense Fog," *IEEE Photonics Technol. Lett.*, vol. 27, no. 4, pp. 383–386, Feb. 2015, doi: 10.1109/LPT.2014.2375795.
- [14] G. D. Forney and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, Art. no. 6, Oct. 1998, doi: 10.1109/18.720542.
- [15] F. R. Kschischang and S. Pasupathy, "Optimal nonuniform signaling for Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, Art. no. 3, May 1993, doi: 10.1109/18.256499.
- [16] P. Schulte and G. Böcherer, "Constant Composition Distribution Matching," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, Art. no. 1, Jan. 2016, doi: 10.1109/TIT.2015.2499181.
- [17] D. Piloni, F. Forghieri, and G. Bosco, "Maximization of the Achievable Mutual Information using Probabilistically Shaped Squared-QAM Constellations," in *Optical Fiber Communication Conference*, Los Angeles, California, 2017, p. W2A.57. doi: 10.1364/OFC.2017.W2A.57.

# Experimental Investigation of Information Bit Scrambling for Physical-Layer Security in Coherent Fiber-Optic Systems

Carsten Schmidt-Langhorst<sup>1</sup>, Johannes Pfeiffer<sup>2</sup>, Robert Elschner<sup>1</sup>,  
Robert Emmerich<sup>1</sup>, Fabian Chowanek<sup>1</sup>, In-Ho Baek<sup>1</sup>, Robert F.H. Fischer<sup>2</sup>, Colja Schubert<sup>1</sup>

<sup>1</sup>Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institute, Berlin, Germany

<sup>2</sup>Institute of Communications Engineering, Ulm University, Ulm, Germany

Email: carsten.schmidt-langhorst@hhi.fraunhofer.de

**Abstract**—In fiber-optic transmission systems, security against eavesdroppers is of increasing importance. Resource-efficient methods of physical-layer security (PLS) are particularly attractive, as they allow for secure optical transmission without requiring secret cryptographic keys known to both communication parties. Combining conventional forward-error correction (FEC) coding with information bit scrambling is a way to achieve PLS. For security, a linear feed-forward shift register is employed at the transmitter side as information bit scrambler, which necessitates the use of a corresponding (self-synchronized) linear feedback shift register to recover the data at the receiver side. Therefore, residual post-FEC errors in the eavesdropped signal create an error avalanche in the eavesdropper's descrambler.

We experimentally assess the achievable security gaps by carrying out back-to-back system experiments using a polarization-division multiplexing (PDM) digital-coherent optical system setup operated at a symbol rate of 64 GBd with offline processing. The security gaps of the scrambling method are compared for quadrature phase shift keying (PDM-04QAM), 16ary quadrature amplitude modulation (PDM-16QAM) and PDM-64QAM using conventional error-correcting low-density parity-check (LDPC) codes. This results in up to 640 Gb/s net secure data rates.

Compared to our previous experimental investigations of special punctured LDPC codes (targeting both, reliability and security), the combination of conventional LDPC codes with information bit scrambling yields significantly improved (reduced) security gaps.

**Keywords**— *Fiber optics, Optical communications, Physical-layer security, Security gap, Information scrambling*

## I. INTRODUCTION

Resource-efficient protection against eavesdroppers is an increasing concern in many communication systems. This includes fiber-optical communications, where wiretapping may be realized without disrupting the service, e.g., via fiber bending [1].

Usually, security and confidentiality are primarily addressed on the upper layers of the Open Systems Interconnection (OSI) model and rely on the philosophy of computational security. However, secure message transmissions without the need for secret cryptographic keys can be achieved by techniques of the physical layer, i.e., by

applying suitable coding and modulation schemes, cf. [2]. A prerequisite for the resource-efficient methods of physical-layer security (PLS) is some advantage of the legitimate receiver (Bob) over potential eavesdroppers (Eve), e.g., better channel conditions in terms of a higher optical signal-to-noise ratio (OSNR).

An eavesdropping scenario can be represented and analyzed using a wiretap channel model [3]. Both Bob and Eve face the same decoding problem. If identical receivers are employed, their operating points lie on the same bit error ratio (BER) curve at different OSNRs. So, coding schemes are required, which yield an almost error free transmission for Bob and, at the same time, result in a high BER for Eve, even if Eve operates only at a slightly smaller OSNR than Bob.

The security gap metric introduced in [4] enables a security assessment of practical coding and modulation schemes based on their BER curve. The metric directly yields the required OSNR advantage of Bob over Eve such that some BER requirements are fulfilled. The two involved parameters are Bob's maximum BER, such that the transmission is reliable, i.e., close to zero, and  $BER_{Eve,min}$ , a BER close to 0.5, which Eve must exceed in order to have a secure transmission. Via the BER curve, they directly translate to the corresponding OSNRs, whose ratio (difference in dB scale) is defined as security gap  $S_g$ .

In this paper, we experimentally assess practically achievable security gaps using shift-register based information bit scrambling in combination with low-density parity-check (LDPC) channel codes in an optical digital-coherent back-to-back system experiment. We use bit-interleaved coded modulation (BICM) to generate polarization-division multiplexed quadrature-amplitude modulated signals carrying two (PDM-04QAM), four (PDM-16QAM), and six (PDM-64QAM) bits per symbol per polarization at a symbol rate of 64 GBd. Offline digital signal processing (DSP) is used to evaluate coded message transmissions, cf. [5].

Compared to previously examined punctured LDPC coding schemes, cf. [6], the employed scrambling scheme yields both, greatly reduced security gaps and higher secure transmission rates for similar set-ups. In addition, operating points of the legitimate receiver are more favorable, in terms of smaller implementation penalties and, thus, OSNRs.

We consider the following three cases for which security gaps are derived: (1) as a reference, both Bob and Eve operate with ideal receivers over the additive white Gaussian noise (AWGN) channel, (2) as a worst-case assessment, Bob

This work has been carried out within the framework of the CELTIC EUREKA project AI-NET-PROTECT (Project ID C2019/3-4) and was funded by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF) under contracts 16KIS1282 and 16KIS1287.

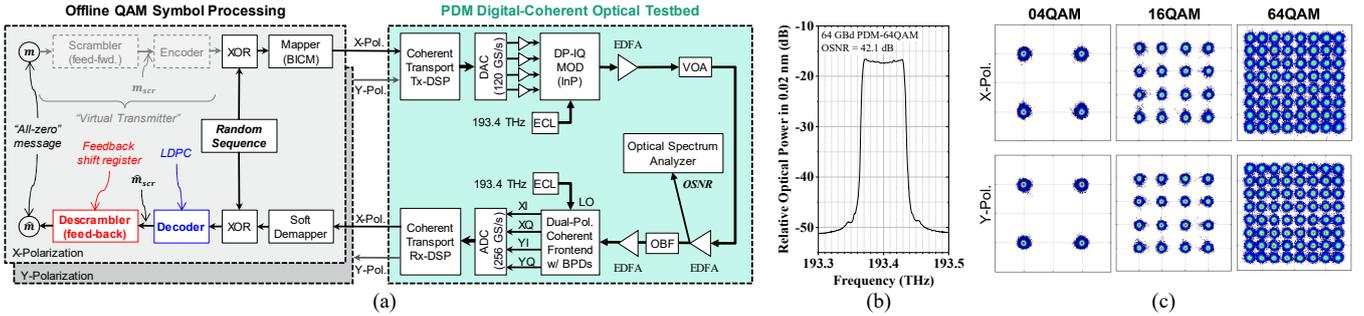


Fig. 1. a) Schematic depiction of experimental set-up and used offline DSP; b) Optical spectrum at the output of the transmitter for 64 GBd PDM-64QAM; c) Constellation diagrams after coherent Rx-DSP (i.e., soft demapper input) for 64 GBd PDM-04QAM, PDM-16QAM and PDM-64QAM (OSNR  $\sim$ 40 dB).

operates with a real-world coherent receiver while Eve is assumed to operate an ideal receiver over the AWGN channel, and (3) both Bob and Eve employ the same state-of-the-art real-world coherent receiver.

This paper is organized as follows: Sec. II gives an overview of the concept of information bit scrambling and its SR realizations. Our experimental set-up consisting of a coherent optical testbed and offline digital signal processing (DSP) is presented in Sec. III. In Sec. IV, experimental results, the related security gap evaluation, and a discussion on optical polarization impacts are given. This paper closes with a brief summary and conclusions in Sec. V.

## II. INFORMATION BIT SCRAMBLING

In order to achieve a sharp transition between operating points of high and small BERs, information bit scrambling can be applied, cf. [7],[8],[9]. To that end, prior to channel encoding at the transmitter, secret binary messages  $m$  are scrambled into messages  $m_{scr}$  of the same length. Consequently, any receiver, which wants to recover  $m$  must apply descrambling on the output of its channel decoder  $\hat{m}_{scr}$  to get a message estimate  $\hat{m}$ .

Bob, who operates at a high OSNR, can usually recover  $\hat{m}_{scr}$  error freely and, consequently,  $\hat{m}$  is also free of errors. In contrast, at a slightly smaller OSNR, Eve will typically experience some errors in  $\hat{m}_{scr}$ . When descrambling, error propagation leads to a large number of errors in  $\hat{m}$ . The structure of the descrambler, which in turn must match to the transmitter's scrambler, mainly determines the final error rate.

In the following, the employed scrambling scheme is based on efficiently implementable linear feedback shift registers (SRs). As shown in [9], it is favorable to force receivers to employ a purely recursive SR as descrambler by using a non-recursive feed-forward SR as scrambler. This configuration enables a high error propagation even for SRs of small sizes. To prevent unconstrained error propagation, which would severely affect the legitimate receiver, both scrambler and descrambler are reset to a fixed state (all zeros) after each information word.

For our evaluations the scrambler is described by the difference equation  $m_{scr}[k] = m[k] + m[k-14] + m[k-15]$ , where all quantities are binary, addition corresponds to the XOR operation, and  $k$  denotes the discrete time index.

As for all other parameters of the communication set-up, information about the scrambler is perfectly known to all receivers. There, descrambling can be achieved by a recursive SR, which realizes  $\hat{m}[k] = \hat{m}_{scr}[k] + \hat{m}[k-14] + \hat{m}[k-15]$ .

## III. EXPERIMENTAL SET-UP

The experimental investigations were carried out using a coherent optical system testbed with offline digital signal processing (DSP) as schematically depicted in Fig. 1(a).

### A. PDM digital-coherent optical testbed

In the coherent optical testbed, Alice's optical data signal (PDM-QAM symbols) is modulated onto an optical carrier from an external cavity laser (ECL, 193.4 THz optical frequency,  $<100$  kHz linewidth) by an indium phosphide (InP) dual-polarization in-phase and quadrature modulator (DP-IQ MOD) driven by four integrated electrical driver amplifiers (40 GHz bandwidth). The analog electrical drive signals are generated by a four-channel digital-to-analog converter (DAC, 120 GS/s, 45 GHz bandwidth). The DAC waveforms were pre-processed offline using a coherent transport transmitter-side DSP (Tx-DSP).

In the coherent Tx-DSP, 448 header symbols are periodically inserted (every  $32 \cdot 320$  symbols) into a sequence of 64 GBd PDM-QAM data symbols coming from the QAM symbol processing DSP (see below). The header symbols include constant-amplitude zero-autocorrelation (CAZAC-64) training sequences. Next, a root-raised cosine pulse shape is applied (roll-off 0.1) and linear pre-distortion of the transmitter components is applied. Finally, the waveforms are resampled to the DAC sampling rate (120 GS/s) before storage in the DAC memory.

Fig. 1(b) exemplarily shows the optical transmitter output spectrum for the case of 64 GBd PDM-64QAM at an OSNR of 42.1 dB and illustrates the applied Nyquist pulse shape and pre-distortion.

At the receive side, Bob uses an optically pre-amplified digital coherent receiver. Here, the received optical data signal is first optically amplified using two Erbium-doped optical amplifiers (EDFA) with an intermediate optical bandpass filter (OBF) to remove out-of-band amplified spontaneous emission noise. Next, it is mixed with the optical local oscillator (LO, generated from a separate ECL at 193.4 THz with  $<100$  kHz linewidth) in a dual-polarization coherent frontend comprising a dual-polarization quadrature optical hybrid and four balanced photodetectors (BPD, 100 GHz bandwidth). The four analog electrical output signals of the balanced detectors are digitized by a four-channel analog-to-digital converter (ADC, 256 GS/s, 113 GHz bandwidth). The digitized received waveforms are then post-processed offline using the coherent transport receiver-side DSP (Rx-DSP).

The state-of-the-art coherent transport Rx-DSP comprises front-end corrections, resampling to  $2$  Sa/Sym, training-aided

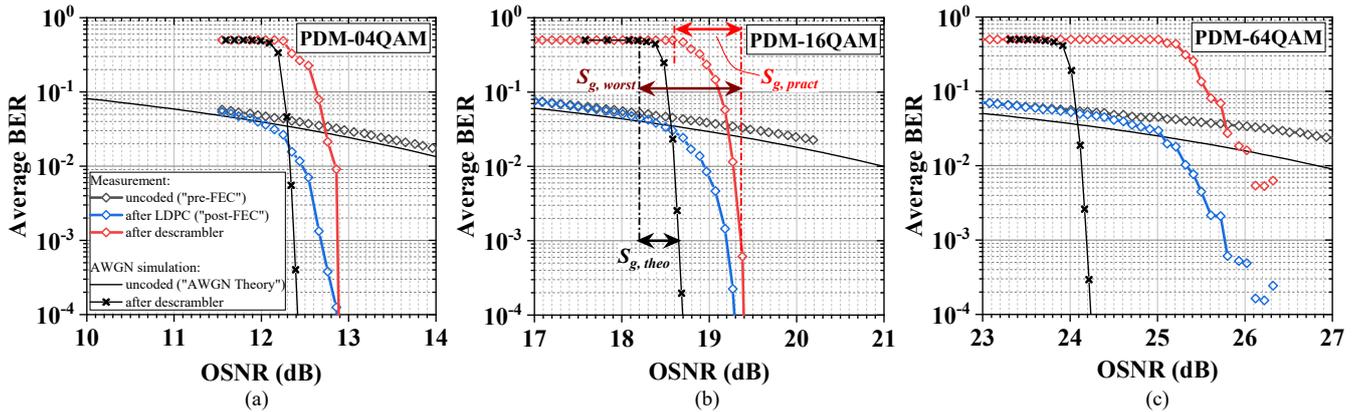


Fig. 2. Measured average BER vs. OSNR for (a) PDM-04QAM, (b) PDM-16QAM, (c) PDM-64QAM. The legend in (a) applies to all plots. The different security gap metrics used throughout this work are exemplarily illustrated in (b) and apply to all modulation formats. Please note, that in (c) the evaluated measurements yielded no bit error after LDPC and after descrambler for OSNR > 26.4 dB.

frame synchronization, frequency offset estimation and channel estimation, and T/2-spaced 2x2 MIMO frequency-domain equalization including polarization separation and downsampling to 1 Sa/Sym. This is followed by a carrier-phase estimation using the blind phase search algorithm and T-spaced adaptive equalization using a decision-directed, real-valued 4x4 equalizer. The resulting PDM-QAM symbols are then further post-processed in the QAM symbol DSP (see below).

In the experiments, a variable optical attenuator (VOA) was used to emulate the link loss between Alice and Bob. The resulting received optical signal-to-noise ratio (OSNR, in 0.1 nm) was measured using an optical spectrum analyzer which was placed after the EDFA at the input of the receiver.

Received constellation diagrams after the coherent Rx-DSP are shown in Fig. 1(c) for all three investigated modulation formats, namely PDM-04QAM, PDM-16QAM, and PDM-64QAM, at maximum OSNR.

### B. Offline QAM symbol processing

The PDM-QAM symbols to be transmitted are created separately per polarization from random binary information bit sequences by offline pre-processing as depicted in the left part of Fig. 1(a). Bit-interleaved coded modulation (BICM) in combination with Gray labeling is applied. Correspondingly, the received PDM-QAM symbols are post-processed by polarization to get the BER performance of the examined coding scheme. A soft-output demapper derives log-likelihood ratios for all communicated bits based on the utilized signal constellation.

In the transmission experiments, a random bit sequence is communicated rather than actual codewords of the examined rate-5/6 (20 % overhead) length-64800 LDPC code, cf. [10]. To assess the performance of this channel code in combination with information bit scrambling, the offline evaluation approach from [5] is employed, which outlines the processing of measurement data w.r.t. various coding schemes and has also been used in [6], [11].

As depicted in Fig. 1(a), a side-stream scrambler is inserted in the transmit path between the encoder and mapper and in the receive path between the soft demapper and LDPC decoder. In the virtual transmitter, a random binary message  $m$  is first scrambled into  $m_{scr}$  and then encoded into code words  $c_{virt}$ , which are subsequently additively scrambled (XOR) into the actual transmitted words  $c_{act}$ . In the receiver,

In the receiver, the side-stream scrambling has to be reversed. The operation is performed on the log-likelihood ratios, where a bit flip corresponds to a sign change. The processing is split into blocks of length  $N = 64800$ , corresponding to the block length of the used LDPC code.

The channel decoder can then process the log-likelihood ratios and yield an estimate of the scrambled message  $\hat{m}_{scr}$ . In case of successful decoding,  $\hat{m}_{scr}$  is the information word corresponding to the codeword  $c_{virt}$ .

$\hat{m}_{scr}$  is then passed to the descrambler to obtain  $\hat{m}$ , an estimate of the communicated message. Comparing  $\hat{m}$  and  $m$  gives the number of bit errors in that information word.

This process is repeated for all words within a waveform and for all recorded OSNRs to derive BER curves as seen in Fig. 2. As usual, the BER performance of both polarizations is averaged into a single BER curve.

## IV. EXPERIMENTAL RESULTS

In this section, we report on the experimentally obtained bit-error ratio performance and quantitatively assess the results using the security gap metric. Finally, we address polarization-related practical aspects when evaluating security gaps in polarization-division multiplexed systems.

### A. Bit-error ratio performance

The average BER performance of all investigated formats as a function of the OSNR (in 0.1 nm) is shown in Fig. 2. Each plot shows the measured “uncoded” BER before the LDPC decoder, which is commonly referred to as “pre-FEC BER” (black diamonds). For reference, the theoretical performance for uncoded BER over an AWGN channel is also shown (black line).

Moreover, the plots contain the BER of the received message bits after descrambling (red diamonds). For comparison, the simulated BER performance after descrambling assuming an AWGN channel and ideal optical components is shown (black crosses).

For illustration purposes, we also plot the BER after the LDPC decoder, i.e., the BER performance of a conventional transceiver without scrambling (blue diamonds). It can be seen that the error correction threshold of the applied LDPC code is about  $3 \cdot 10^{-2}$ , as expected for the applied LDPC code rate of 5/6 corresponding to 20% overhead.

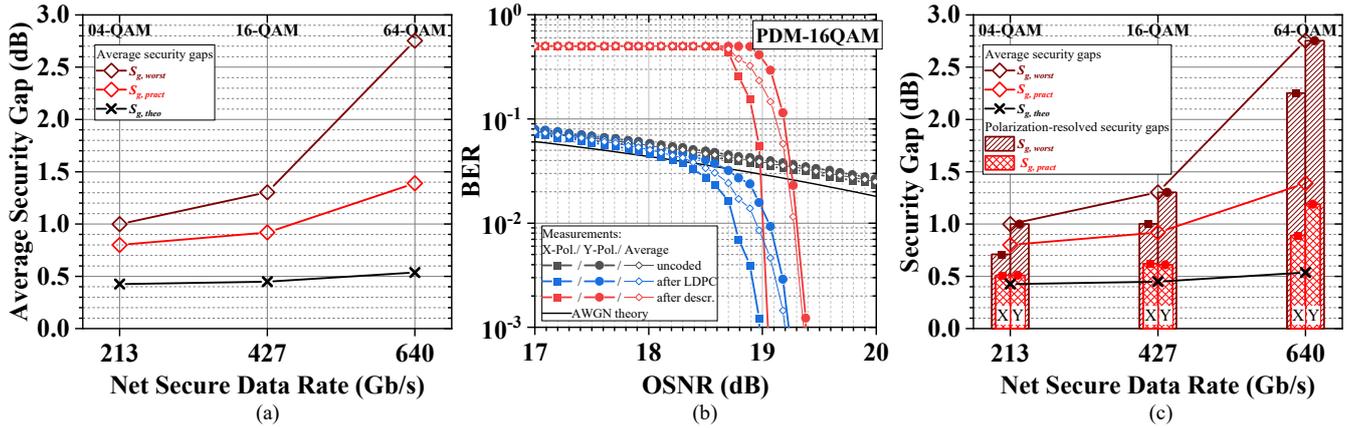


Fig. 3. (a) Average security gap vs. net secure data rate (accounting for code rate 5/6 and symbol rate 64 GBd) of the investigated modulation formats; (b) Polarization-resolved BER performance of 64 GBd PDM-16QAM; (c) Polarization-resolved security gaps for all investigated modulation formats.

Comparing the measured pre-FEC BER with the AWGN performance reveals the so-called “implementation penalty” of the practically implemented transceiver setup. At a BER of  $3 \cdot 10^{-2}$  we find implementation penalties of about 0.4 / 0.5 / 1.8 dB for PDM-04QAM, PDM-16QAM and PDM-64QAM, respectively. The increasing implementation penalties with increasing cardinality is a typical effect in practical coherent optical systems arising from nonlinear distortions in the creation and reception of multilevel optical signals in each quadrature.

Comparing the measured BER after descrambling with the numerical simulation of the BER after descrambling we find an increasing OSNR penalty with increasing cardinality of the modulation format. This is a direct consequence of the practical implementation penalty discussed above.

### B. Security gap evaluation

We use the security gap metric to quantify the security levels obtained for the different modulation formats. For our evaluations, we define the required maximum BER for Bob to be  $10^{-3}$ , while the required minimum BER for Eve is 0.49.

As an extension to the original definition of security gap, we here consider three different variants of security gaps, depending on different assumptions about the used hardware for the legitimate receiver Bob and the eavesdropper Eve:

- Theoretical security gap,  $S_{g,theo}$ :

Here, both Bob and Eve are assumed ideal, i.e., both operate on the numerically simulated, theoretical AWGN BER performance. This is equivalent to the security gap  $S_g$  typically used in the literature, cf. [4], [6], [9], [11].

- Practical security gap,  $S_{g,pract}$ :

Here, Bob and Eve use the same practical receiver implementation, i.e., both operate on the measured BER curve and are impacted by the same practical implementation penalty.

- Worst-case security gap,  $S_{g,worst}$ :

Here, Bob uses the practical receiver implementation and operates on the measured BER performance, while Eve is assumed to use an ideal receiver and to operate at the AWGN performance, cf. [6], [9].

Due to the practical implementation penalties discussed above, it is expected that

$$S_{g,theo} < S_{g,pract} < S_{g,worst}.$$

Fig. 3(a) shows the derived security gaps for all investigated formats as a function of the net secure data rate. As the descrambler does not alter the data rate, the secure net data rate is equal to the “conventional” net data rate (i.e., after FEC/LDPC) of the coherent transmission system. At a symbol rate of 64 GBd and an LDPC code rate of 5/6, this results in net secure data rates of 213 / 427 / 640 Gb/s for PDM-04QAM, PDM-16QAM, and PDM-64QAM, respectively.

From Fig. 3(a), it can be seen that the theoretical security gaps  $S_{g,theo}$  are about 0.5 dB for all three constellations, with a slight increase towards higher cardinality constellations due to the used BICM approach (in agreement with [9]).

The measured practical security gaps  $S_{g,pract}$ , however, are found to be 0.8 dB (213 Gb/s PDM-16QAM), 0.9 dB (427 Gb/s PDM-16QAM), and 1.4 dB (640 Gb/s PDM-64QAM). They exhibit an offset w.r.t. the theoretical security gaps due to the aforementioned implementation penalties, which “flatten” the waterfall region of the LDPC code performance and thereby increase the gap between Bob’s and Eve’s operating points. This offset increases in particular for PDM-64QAM, in agreement with the observed increased implementation penalty of this format.

Finally, the worst-case security gaps  $S_{g,worst}$  are found to be 1.0 dB (PDM-QPSK), 1.3 dB (PDM-16QAM), and 2.8 dB (PDM-64QAM). Here, both, the shift of the measured performance towards larger required OSNR for Bob as well as the flattening of the waterfall region add up in the security gap calculation with Eve being assumed ideal and operating on the AWGN performance.

Nevertheless, when comparing those obtained worst-case security gaps (on the order of 1...3 dB) to earlier experimental investigations of a PLS approach based on punctured LDPC codes (with security gaps on the order of 6...11 dB, cf. [6], [9]), we find that the PLS scheme of information bit scrambling investigated in this work yields significantly reduced security gaps. The main reasons for this improvement are the increase in code length by one order of magnitude and the changed coding scheme (to conventional LPDC). Both lead to steeper BER curves and, thereby, a smaller required OSNR for Bob’s operating point. Additionally, the used LPDC code offers a higher pre-FEC BER correction threshold. Thus, the impact of the practical implementation-induced flattening of the pre-FEC BER curve of Bob is reduced, which especially improves the

security gaps for larger signal constellations like PDM-64QAM.

Moreover, from a practical perspective, the employed linear de-scrambling after the LDPC decoder maintains the error correction performance of the employed standard LDPC code, i.e., it preserves the required OSNR for error-free reception at the legitimate receiver Bob, even with descrambling. This is different from the previously investigated case of using a punctured LDPC code for both, reliability and security, in which Bob's operating was shifted towards larger OSNR, i.e., an additional margin had to be included.

### C. Sensitivity to polarization-dependent BER performance

In a practical realization of a coherent optical transmission system, polarization-dependent component imperfections at the transmitter (e.g., modulator PDL, driver amplitudes) as well as at the receiver (e.g., frontend PDL, photodiode sensitivities) typically cause slight performance differences among the polarization tributaries of the optical PDM data signal. Such imbalances may also be introduced and/or enhanced by polarization dependent components in the link (e.g., EDFAs, filters) or even by the transmission fiber itself.

To illustrate the impact of such polarization imbalances, Fig. 3(b) shows the BER performance per polarization (X-polarization: squares, Y-polarization: circles) for the PDM-16QAM case. It can be seen that in our experimental realization the pre-FEC BER performance is very similar for both polarization tributaries (black curves). Nevertheless, the steep waterfall of the employed LDPC code results in a noticeable spread of the BER performance after decoding (blue curves) and consequently also after descrambling (red curves). For comparison, the average BER is plotted as well (diamonds).

Note that, when averaging the BER curves, the worse tributary determines the required OSNR for Bob ( $10^{-3}$ , in the waterfall region of the BER curve) while the better tributary dominates the allowed OSNR for Eve (high BER region).

To further illustrate the impact of the individual polarization performance, Fig. 3(c) shows the security gaps when evaluated per polarization (X-polarization: squares; Y-polarization: circles) compared to when evaluated from the average BER (diamonds, same as Fig. 3(a)). It can be seen that for PDM-04QAM (213 Gb/s net secure rate), the experimentally obtained per-polarization practical security gap (red bars and symbols) is about 0.5 dB for both polarizations which is only slightly above the theoretical security gap (0.4 dB). This is due to the fact that the experimental PDM-04QAM realization has a low error floor and thus the pre-FEC performance is only minimally flattened w.r.t. the AWGN theory, resulting in an LDPC waterfall which is almost as steep as in the theoretical performance. However, the *average* practical security gap becomes 0.8 dB due to the difference in implementation penalty ("horizontal" shift of BER curves in Fig. 3(b)) for the two polarizations in combination with the aforementioned different dominances of the better/worse polarizations for the operating points of Bob and Eve.

The same relations are also observed for PDM-16QAM, however, at slightly increased security gaps of 0.6 dB and 1.4 dB per polarization and on average, respectively. For PDM-64QAM, the per-polarization practical security gaps

are found to be 0.9 dB and 1.2 dB for X- and Y-polarization, respectively. This increase and spread is due to the higher pre-FEC error floor of the PDM-64QAM implementation, which noticeably flattens the post-LDPC waterfall. The average practical security gap is 1.4 dB and is again dominated by the Y-polarization in our experimental implementation.

Similar trends can also be observed for the worst case security gaps (dark brown bars and symbols). Here, the polarization-dependent horizontal shift of the waterfall region (and thus the operating point of Bob in the worse polarization) directly translates into a polarization dependent security gap (because Eve's operating point is polarization-insensitive in the AWGN simulation).

In summary, the security gap metric is rather sensitive to minor per-polarization implementation differences and resulting differences in per-polarization pre-FEC BER, especially if long LDPC codes with steep waterfall characteristics are employed.

## V. CONCLUSION

While in today's optical transmission systems reliability is achieved using soft-decision error correcting codes, tap-proof security can be seamlessly and resource-efficiently (in terms of ASIC implementation) added by information bit scrambling using linear shift registers.

Using a polarization division multiplexing, digital coherent optical transmission testbed operating at a symbol rate of 64 GBd, we evaluated the achievable security gaps for PDM-04QAM, PDM-16QAM and PDM-64QAM. We found average worst-case security gaps (Bob uses the practical receiver implementation and Eve is assumed ideal) between 1 dB for PDM-04QAM (213 Gb/s net secure data rate) and 2.8 dB for PDM-64QAM (640 Gb/s net secure data rate). Considering the case that Eve also uses a practical receiver implementation, the resulting practical security gaps are reduced to below 1 dB for PDM-04QAM and PDM-16QAM and to 1.4 dB for PDM-64QAM.

Detailed investigations revealed a noticeable sensitivity of the security gap metrics towards even small polarization imbalances in the practical implementation. The typically considered average BER and resulting average security gaps, however, serve as a worst case boundary and thus are still useful to assess the security level of different PLS schemes.

Compared to our previous experimental investigations of special punctured LDPC codes (targeting both, reliability and security), the combination of conventional LDPC codes with information bit scrambling as presented in this work yields significantly improved (reduced) security gaps. Moreover, from a practical perspective, the employed linear de-scrambling after the LDPC decoder maintains the required OSNR for error-free reception at the legitimate receiver, i.e., no additional margin has to be allocated.

## REFERENCES

- [1] M. Zafar Iqbal, H. Fathallah, N. Belhadj, "Optical fiber tapping: Methods and precautions," 8th International Conference on High-capacity Optical Networks and Emerging Technologies, Dec. 2011, pp. 164–168.
- [2] M. Bloch, J. Barros, *Physical-Layer Security: From Information Theory to Security Engineering*, Cambridge University Press, 2011.
- [3] A. Wyner, "The Wire-Tap Channel," The Bell System Technical Journal, vol. 54, no. 8, pp. 1355–1387, Oct. 1975.

- [4] D. Klinc, J. Ha, S.W. McLaughlin, J. Barros, B.J. Kwak, “LDPC Codes for the Gaussian Wiretap Channel,” *IEEE Trans. on Information Forensics and Security*, vol. 6, no. 3, pp. 532–540, Sep. 2011.
- [5] L. Schmalen, F. Buchali, A. Leven, S. ten Brink, “A Generic Tool for Assessing the Soft-FEC Performance in Optical Transmission Experiments,” *IEEE Phot. Techn. Lett.*, vol. 24, no. 1, pp. 40–42, Jan. 2012.
- [6] C. Schmidt-Langhorst, J. Pfeiffer, R. Elschner, F. Frey, R. Emmerich, R. F.H. Fischer, C. Schubert, “Experimental Investigation of Security Gaps when using Punctured LDPC Coding in Coherent Systems up to 64-GBd DP-64QAM,” in 20. ITG-Symposium on Photonic Networks, May 2019, pp. 55–57.
- [7] K. Guan, J. Cho, P. J. Winzer, “A Computationally Efficient Shift-Register Based Information Scrambling Approach to Physical Layer Security in MIMO-SDM Systems,” in *Optical Fiber Communications Conference and Exhibition (OFC)*, Mar. 2016, pp. 1–3.
- [8] J. Cho, K. Guan, S. Chandrasekhar, P. Winzer, “Experimental Demonstration of Physical-Layer Security in a Fiber-Optic Link by Information Scrambling,” in 42. *European Conference on Optical Communication (ECOC)*, Sep. 2016, pp. 1–3.
- [9] J. Pfeiffer, R. F.H. Fischer, “Combining Scrambling and Multilevel Coding for Physical-Layer Security,” in 22. *ITG-Symposium on Photonic Networks*, May 2021, pp. 43–50.
- [10] ETSI EN 302 307 V1.1.1: Digital Video Broadcasting (DVB); Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications (DVB-S2). Standard, ETSI, EBU, March 2005.
- [11] J. Pfeiffer, C. Schmidt-Langhorst, R. Elschner, F. Frey, R. Emmerich, C. Schubert, R. F.H. Fischer, “Security Gap Investigation of Multilevel Coding in Coherent Fiber-Optical Systems,” in 21. *ITG-Symposium on Photonic Networks*, Nov. 2020, pp. 40–46.

# Experimental validation of exact component parameter agnostic QoT estimation using spectral data-driven LSTM over a 265.2 km SSMF link

Lars E. Kruse  
*Chair of Communications*  
*Christian-Albrechts-University of Kiel*  
 Kiel, Germany  
 lars.kruse@tf.uni-kiel.de

Stephan Pachnicke  
*Chair of Communications*  
*Christian-Albrechts-University of Kiel*  
 Kiel, Germany  
 stephan.pachnicke@tf.uni-kiel.de

**Abstract**— The ever-growing demand for capacity necessitates optical networks to undergo enormous changes which will lead to the increase in complexity of future networks. Since the beginning of optical transmission networks, the insurance of services to provide the best possible connection for the customer (Quality of service, QoS) has always been an important topic for network providers. Traditionally, networks were conservative with high margins. However, high margins lead to wasted capacity. Due to this, the optimization of new and established connections is of increasing importance. Quality of transmission (QoT) estimation enables network operators to respond proactively to performance degradations and to lower the design margins. Machine learning-based QoT estimation has received significant attention from the community, since it can become an alternative to numerical and analytical approaches, which suffer from high computational complexity or inaccuracy, respectively.

In this paper, we extend our work [1] by validating the proposed QoT estimation tool based on a spectral data-driven LSTM/NN-based machine learning algorithm by testing the algorithm, trained with simulation data, on an experimental dataset that contains data from a DP-QPSK or DP-16-QAM modulated WDM transmission over 3 spans of 88.4 km SSMF with different channel assignments of 9 channels in a 37.5 and 50 GHz ITU-grid operating at 32 Gbaud. The estimator shows an MAE of 0.507 dB and an  $R^2$ -score of 0.897.

**Keywords**— *machine learning, quality of transmission estimation, agnostic network scenarios*

## I. INTRODUCTION

Over the past decade, optical networks have changed enormously to meet the ever-increasing demand for capacity. In this regard, networks have become more sophisticated, configurable, and adaptable. New flexible add-drop multiplexers enable more versatile network operation as well as the implementation of flexible frequency grids. Due to the significant increase in network complexity, monitoring and optimizing performance is of increasing importance. An accurate estimation of the quality of transmission (QoT) allows to maximize capacity and may enable full self-management of the networks in the future. While QoT estimation in optical networks has been around for a long time, the changes mentioned above, as well as the possibility of using data-driven, autonomous artificial intelligence (AI) techniques have renewed interest in reliable QoT estimation.

In recent years, much effort has been invested in the development of analytical [2], machine learning-based techniques [3,4], and hybrid approaches [5,6,7] to evaluate the performance of a particular lightpath in the system, based on its generalized SNR (gSNR) [4]. All of the aforementioned works require detailed knowledge of the necessary component parameters for QoT estimation. However, in practice, in many cases, the network operator does not have precise knowledge of all parameters of the deployed network elements (e.g., the real noise figure of all amplifiers, exact fiber parameters, etc.); we refer to such a scenario as exact component parameter agnostic or agnostic for short. A GN model-based estimator, due to its analytical nature, would require exact knowledge of the parameters or otherwise would require high margins, while machine learning-based QoT estimation has been shown to be fast and accurate [3,4,5].

This paper is an extension of our work from the OFC 2022 [1], which was a purely theoretical study, and is organized as follows. We present previous work on the subject in Section II. In Section III the QoT estimator is explained in detail. Section IV contains the experimental results, and in Section V we conclude the paper.

## II. PREVIOUS WORK

Maximizing the available capacity of a deployed network is the main goal of operators, which are willing to improve their network efficiency given the highly increasing volume and resource-constrained long-haul fiber links [9]. Accurate QoT estimation is fundamental for achieving this goal.

A machine-learning (ML) physical layer model trained with physical layer parameters and an ML model using monitoring metrics combined with analytically generated link-based features was found to achieve very good accuracy with a relatively small training database [5]. ML-based methods have also been explored for predicting performance metrics such as the bit-error-rate (BER), signal-to-noise ratio (SNR), or Q-Factor for deployed lightpaths [4]. Performance prediction would allow network operators to respond proactively to performance degradations or potential failures in optical networks.

The optical signal-to-noise ratio (OSNR) is a primary optical link performance metric. Monitoring the OSNR enables network operators to validate the expected performance from the network planning stage since the aforementioned metrics are closely connected to the OSNR.

Furthermore, it is possible to monitor the health of the optical system permitting proactive maintenance and to optimize the link capacity and/or margin [10]. Transceiver impairments limit the informative value of these metrics since the impact of the impairments varies from system to system. Therefore, the generalized-SNR (GSNR) or often GOSNR is defined as the required OSNR to achieve the same BER as in the back-to-back (B2B) case. Generally, the GOSNR captures only the optical impairments including linear noise and nonlinear interference and reflects the BER more accurately. It is defined as

$$GOSNR = \frac{P_{Rx}}{P_{ASE} + P_{NLI}}, \quad (1)$$

where  $P_{Rx}$  is the received power,  $P_{ASE}$  is the power of the linear ASE noise and  $P_{NLI}$  is the power of the nonlinearities.

The state of the art for optical data transport relies on wavelength-division-multiplexing (WDM) transmission and dual-polarization coherent optical technology with the use of software-defined networking (SDN) [11]. The SDN paradigm in optical networks enables the management of each network element. Thus it enables providing a specific lightpath given the current network status and GSNR estimation. However, it has been demonstrated that, in the case of accurate knowledge of the physical layer, an ML-based QoT estimator can give an extremely accurate GSNR computation [11]. This accurate knowledge of the physical layer is not always given in a practical networking scenario. For these exact parameter agnostic cases, the use of spectral data is a promising way to increase the QoT estimation performance [1].

In this work, we will show the result of an experimental validation of our spectral data-driven QoT estimator.

### III. SPECTRAL DATA BASED QoT ESTIMATOR

In [11] optical spectrum analyzers (OSAs) are used to obtain the optical spectrum to identify soft failures caused by filters. The obtained spectrum cannot only be used for failure detection but also reflects the current state of the connection. Certain values extracted from the spectrum at selected relevant nodes in a network can increase the estimation accuracy of a ML-based QoT estimator as we have shown in our previous work in [1].

#### A. Simulation Setup

For the data generation, we set up a simulation setup as depicted in Fig. 1. The setup consists of a database of the

TABLE I. VARIED SIMULATION PARAMETERS

Parameter	Symbol	Mean Value	Standard Deviation
Span lengths	$L_S$	80 km	5 km
EDFA output power	$P_E$	$P_L$	0.5 dBm
EDFA noise figure	$NF$	5 dB	0.5 dB
Attenuation	$\alpha$	0.2 dB/km	0.02 dB/km
Dispersion	$D$	17 ps/(nm·km)	0.2 ps/(nm·km)
Nonlinear coefficient	$P_L$	1.295 (W·km) <sup>-1</sup>	0.05 (W·km) <sup>-1</sup>

obtained feature vectors and the optical transmission system, which defines the environment. A set of  $N_c$  coherent dual-polarization channels,  $\mathbf{c} = [c_1, \dots, c_{N_c}]$ , is to be transmitted over a WDM link with fixed channel spacing and equal launch powers per channel. Since the exact component parameters in an agnostic network are not known, the transmission parameters used for the simulations are calculated using a heuristic approach with a certain mean and standard deviation based on realistic assumptions and margins. The links are analyzed for different modulation formats ( $MF$ ) i.e., QPSK, 8-QAM and 16-QAM. The baudrate ( $b$ ) is changed between 32 and 64 Gbaud. The channel spacing is varied between  $\Delta f$  of 37.5 and 50 GHz for the 32 Gbaud case, and 100 GHz for the 64 Gbaud case. The launch power is varied between -3 and +1 dBm per channel. The channel under test (CUT) of the WDM transmission has a wavelength of 1550 nm. Uncertainties are for example considered in the span lengths ( $L_S$ ) by choosing randomly a length with a mean of 80 km and a standard deviation  $\sigma$  of 5 km. The varied parameters are summarized in Table I. Therefore, the parameters differ per span according to the random distribution. We assume that the transmission is done over  $N_s$  spans, which are composed of standard single-mode fibers (SSMF) followed by an EDFA with a flat gain characteristic in the C-band and an optical spectrum analyzer (OSA) with a resolution of 13 pm (according to specs of commercially available OSAs). The nonlinearities in this dual-polarisation WDM transmission scenario are calculated solving the Manakov-PMD-equations using the split-step Fourier method with a randomly chosen number of waveplates ranging from 50 to 200.  $N_s$  is increasing with every iteration step of the simulation, i.e., in every step we

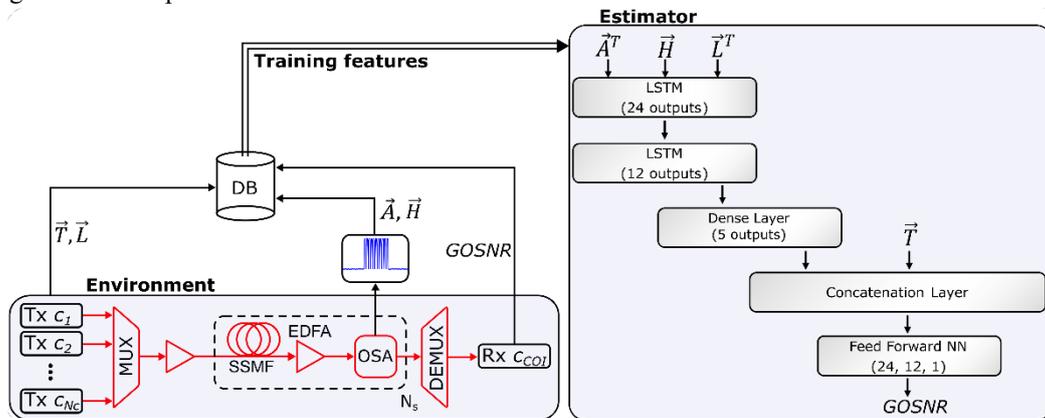


Fig. 1. (left) Simulation setup with a central database (DB) containing the obtained feature vectors and (right) the GOSNR estimator structure realized with LSTM and FF-NN layers.

add one span to the simulation setup until we reach a total of 20 spans. By this, we ensure to represent a network structure in the simulation, since we can interpret the link in different ways. For example, if the link contains 4 spans, we represent a set of links with 0, 1, 2, and 3 intermediate nodes and their different possible distance variations for the distances to and from the intermediate nodes. The GOSNR of the CUT is calculated using equation (1). Here the sum of the linear noise, i.e.  $P_{ASE}$  and the sum of the noise induced by nonlinearities  $P_{NLI}$  is calculated over the deviations of the received constellation points to the ideal ones. This means the GOSNR in the simulation case is equivalent to the inverse EVM, since the back-trace method used in the experiments is not applicable in the simulation cases reasoned by the optimal transceiver characteristics. During the simulation,  $N_c$  is varied from 1 to 11 to represent all possible channel assignment scenarios. Furthermore, different scenarios of adding and dropping channels are considered. Up to a total of 5 neighboring pairs of the CUT are dropped at all the intermediate nodes. For the add scenario on the other hand these neighboring channels are added to the channel configurations in which they are free.

We have set up this scenario for the data generation in our Matlab-based simulation tool. The data extracted from the simulation can be divided into two different categories: The transmission-related features and the features extracted from the spectrum obtained by the OSA. The transmission-related features are composed out of the vector  $\vec{T} = [MF, P_L, \Delta f, b, L]$  with  $L$  being the total link length and the vector  $\vec{L} = [L_{N_S-k, N_S-(k-1)}, \dots, L_{N_S-1, N_S}]$  with  $k$  representing the number of intermediate nodes.  $\vec{L}$  contains the information on the lengths between the intermediate nodes themselves and the end of the link. The spectrum-related features are the vector  $\vec{A} = [A_{N_S-k, N_S-(k-1)}, \dots, A_{N_S-1, N_S}]$  that contains the area under the envelope of the PSD obtained by the OSA, i.e. the total signal power and the vector  $\vec{H} = [H_{c_1}, \dots, H_{c_{11}}]$  which is composed out of the heights of the PSD of the channels  $c$ . In  $\vec{H}$  the channel usage, the influences of uncertainties in the gain, noise figure, attenuation, and the influences of the non-linearities on the power of the channels can be seen. Analyzing the mentioned different modulation formats, launch powers, baudrates and channel configurations result in a dataset with around  $1.5 \cdot 10^6$  feature sets each of which contains up to 53 features. To accommodate more uncertainties in the dataset, the simulation was repeated 10 times, resulting in an even larger dataset of  $15 \cdot 10^6$  features.

## B. Design of the Neural Network

Choosing the optimal machine learning algorithm for the task of QoT estimation is challenging. This is reasoned by the fact that the feature sets, i.e. the feature vectors  $\vec{L}$  and  $\vec{A}$ , are changing their second dimension as the number of considered intermediate nodes is increasing. Besides some support vector machines, there are not many ML algorithms that can operate on a changing number of input features. To solve this problem, we interpret every link as a series of values that can be fed into a recurrent neural network, for example, a long-short term memory network (LSTM). This results in a three-dimensional dataset for example in the shape of  $\vec{L} : [1 \times N_N \times L]$ , where  $N_N$  is the number of intermediate nodes in the link. By changing the shape of the feature set, we ensure that only one step is considered at a time by the LSTM. The same procedure is followed for shaping  $\vec{A}$ .  $\vec{T}$  remains fixed in size so it is fed into the GOSNR estimator in a later stage, as it can be seen on the right of Fig. 1 in which the total estimator structure is depicted. The LSTM layers consist of 24 and 12 outputs, respectively. These parameters were hand-selected and could be further improved using common methods such as grid search or Bayesian hyper-parameter optimization, which however is not the focus of this work. The output is followed by a dense layer, i.e., a fully connected layer with 5 neurons, to match the size of  $\vec{T}$  for the input in the concatenation layer. The feed-forward neural network (FF-NN) consists of 2 hidden layers with ReLU activation functions composed out of 24 and 12 neurons and an output layer with one output which is the GOSNR. Before the training, the obtained dataset is split into 60% training, 20% validation, and 20% test data. The LSTM/FF-NN structure is trained over 1000 epochs and optimized with the Adam optimizer. The trained estimator is tested on the 20% test data from the dataset which results a  $R^2$ -score of 0.9945. To test the robustness of the estimator and to evaluate whether it can be used in reality, the estimator is tested on features obtained through experiments in our lab.

## IV. EXPERIMENTAL VALIDATION

### A. Experimental Setup

The experimental setup is depicted in Fig. 2. The DSP is based on Matlab routines and is executed offline. For the generation of the CUT, at the transmitter side, a PRBS of length  $2^{17}-1$  is generated and mapped to QPSK or 16-QAM. The signal is upsampled from the baudrate of 32 Gbd to the sample rate of the DAC (88 GSa/s) and reshaped using a root-raised cosine filter with a roll-off factor  $\alpha = 0.2$ , resulting in

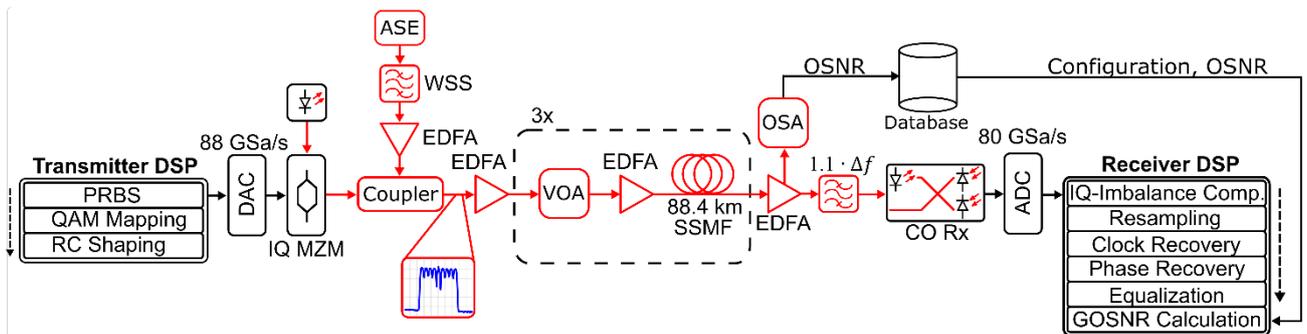


Fig. 2. Experimental setup over 265.2 km of SSMF with variable noise loading for the case of 37.5 GHz WDM channel spacing.

an almost rectangular spectrum. For D/A-conversion, an arbitrary waveform generator running at 84 GSa/s is used. The CUT is generated from an external laser ( $\lambda_{CUT} = 1550.004$  nm) using a DP-IQ modulator that is driven by the DAC via 4 driver amplifiers. The other WDM channels (loaders) are shaped from an ASE noise source comprising an EDFA using a programmable wavelength-shaping filter or short: a WaveShaper (WSS, Finisar WS4000A) with a filter bandwidth of 37.5 or 50 GHz corresponding to the channel spacing. Afterwards the filling channels and the CUT are coupled using a 3dB-coupler. The WSS is configured in such a way, that the CUT and the loading channels have equal power after the coupler. The WDM signal is then amplified using a EDFA to reach the launch powers of -3 to +1 dBm per channel. After the Tx-EDFA, the signal is transmitted over 3 spans consisting of a VOA followed by an EDFA and the standard single-mode fiber (SSMF) with a length of 88.4 km before it is amplified by the receiver side EDFA to compensate for the fiber loss and to guarantee successful detection. The OSNR is measured with an OSA at 0.01 nm resolution and normalized to a reference bandwidth of 0.1 nm. An optical filter with a bandwidth equal to  $1.1 \cdot \Delta f$  is used to filter out the out-of-band noise before detection with a coherent receiver, followed by analog-to-digital conversion using an oscilloscope. An exemplary optical PSD after the transmitter side EDFA is depicted in Fig. 3. In the receiver DSP the signal is first compensated for possible IQ-imbalance before being resampled to  $2 \cdot f_{sym}$ . Afterwards, the signal is synchronized using a constant amplitude zero autocorrelation (CAZAC) sequence. After matched filtering, the signal is equalized using a 2x2 MIMO-equalizer followed by the phase recovery. Then a 4x4 MIMO-equalizer is used to compensate for the polarization-induced interferences. In the end of the DSP chain, different metrics are calculated out of which the GOSNR is the most important for our investigation which we obtained as follows.

The GOSNR includes both the ASE and the nonlinear noise and determines the Q-factor as

$$Q^2 = 20 \cdot \log_{10}(\sqrt{2} \cdot \text{erfc}^{-1}(2 \cdot \text{BER})), \quad (2)$$

and the BER for a given link configuration. The OSNR, however, only captures the in-band ASE noise using the out-

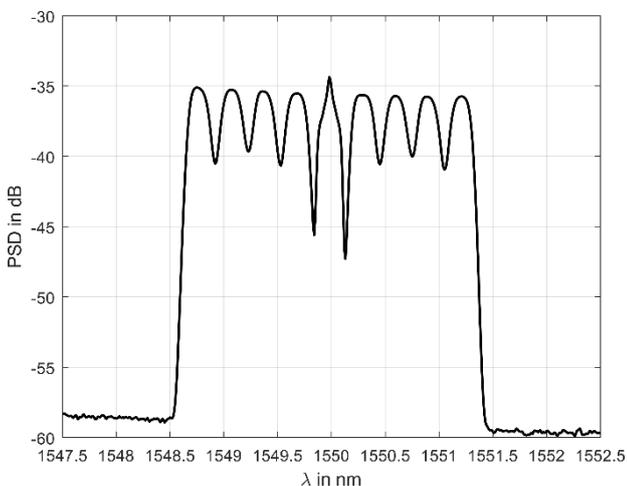


Fig. 3. Spectrum of 16-QAM,  $\Delta f = 37.5$  GHz,  $P_L = +1$  dBm per channel and all loaders active in the back-to-back scenario.

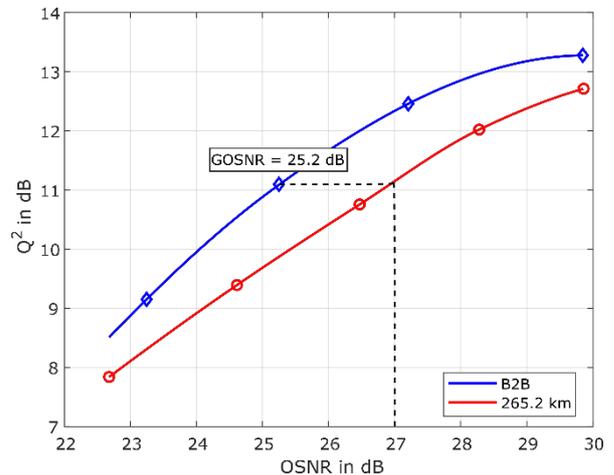


Fig. 4. Backtrace for GOSNR measurement for QPSK,  $\Delta f = 37.5$  GHz and all loaders active at  $P_L = +1$  dBm per channel.

of-band noise floor. The GOSNR of the link is defined as the OSNR required to achieve the same BER in the back-to-back (B2B) case; therefore, the GOSNR is independent of transceiver penalties. In Fig. 3 this backtrace is depicted. The penalty of 2.8 dB in OSNR is therefore added by the fiber nonlinearities within the link. To ensure a faster measurement of the GOSNR, different B2B curves have been collected and have been saved serving as a lookup table for the offline DSP. The curves in Fig. 4 were generated using a VOA before the receiver side EDFA to induce more or less ASE noise to the signal. However, in a real-world scenario, the OSNR changes most significantly with fiber length and the receiver side EDFA always induces roughly the same amount of ASE noise. It can be seen that the maximum GOSNR for QPSK modulation and a channel spacing of  $\Delta f = 37.5$  GHz and all loaders active at  $P_L = +1$  dBm per channel is 27.8 dB. This is caused by the maximum OSNR in the measurements of 29.8 dB (seen in fig.4). A comparison between the measurement and our simulation over 265.2 km SSMF is shown in Fig. 5. With an OSNR between 26.2 and 29.7 dB, the GOSNR values differ only in the range of 0.4 dB. However, the overall GOSNR in the simulations are higher, since not all inaccuracies are modeled, for example the EDFA gain tilt. In general, we can state that our measurement data is very close to the simulation dataset which ensures our ML-algorithm to work.

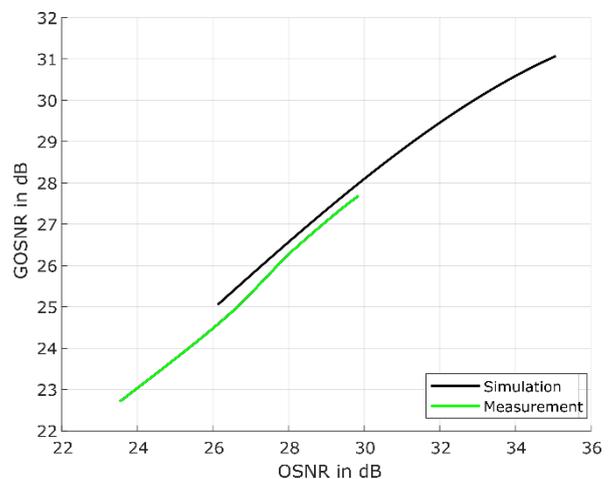


Fig. 5. Comparison of GOSNR simulation and measurement over 265.2 km SSMF for QPSK,  $\Delta f = 37.5$  GHz and all loaders active at  $P_L = +1$  dBm per channel.

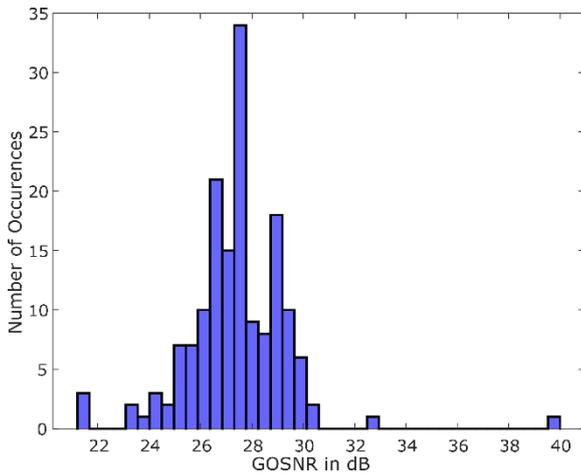


Fig. 6. Occurrence of GOSNR values (distributed in bins of 0.4 dB) of the measurement over 265.2 km SSMF for all configurations.

### B. Results

The estimator trained on the simulation data is tested on the experimentally obtained data. The occurrence of the different GOSNR values can be seen in Fig. 6. It is depicted, that most of the data is in the range of 25 to 29 dB. This is caused by the overall lower OSNR in the measurements. However, due to the normalizing nature of the GOSNR, the values can be compared to the experimental data. The LSTM/NN-hybrid's performance is evaluated on the mean absolute error (MAE) and the  $R^2$ -score. The estimator reached a low MAE of 0.507 dB and an  $R^2$ -score of 0.897. Furthermore, the capability of the ML-algorithm in reacting on experimental data is also shown in the actual against predicted values plot in Fig. 7. The red circles are the actual measured GOSNR values whereas the blue crosses are the estimated values. It is clear to see that there are outliers at the right and left edges of the distribution. The values at the top right of the graph are due to transmission with only one channel, i.e. without the influence of neighboring channels. The outliers at the bottom left, on the other hand, are probably caused by channel interference. These are most likely due to unequal channel power distribution from the EDFA gain tilt and by the non-rectangular filter characteristics. Thus, the interferences and influences of the channels are particularly large in a transmission with only 37.5 GHz channel spacing. In general, however, it can be seen that the distribution of the GOSNR values is quite close to the measured values.

### V. CONCLUSION

In this paper, we demonstrated that our developed estimator enhanced by spectral data obtained by OSAs can also be applied to experimentally obtained data, while being trained with numerical simulations only. The estimation performance with an MAE of 0.507 dB and an  $R^2$ -score of 0.897 is close to the performance achieved in simulations. Furthermore, it has been shown that the estimator can reliably estimate accurate GOSNR values even for span lengths of 88.4 km due to the heuristically distributed training values of the GOSNR.

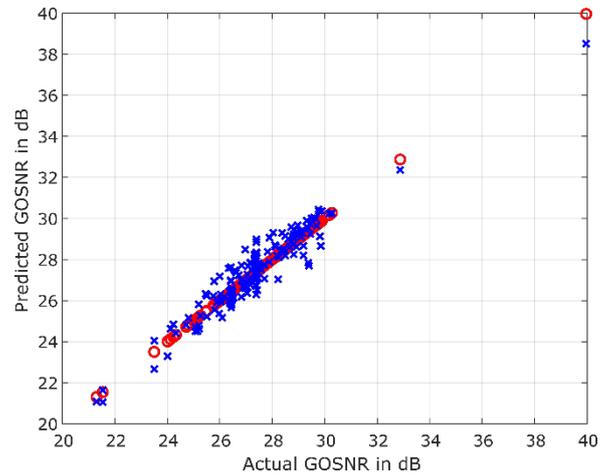


Fig. 7. Predicted GOSNR by the estimator over actual GOSNR for all configurations in the dataset.

### ACKNOWLEDGEMENTS

This work has been performed in the framework of the CELTIC-NEXT project AI-NET-PROTECT (Project ID C2019/3-4), and it is partly funded by the German Federal Ministry of Education and Research (16KIS1284).

### REFERENCES

- [1] L. E. Kruse, S. Pachnicke, "Exact component parameter agnostic QoT estimation using spectral data-driven LSTM in optical networks.", in *Optical Fiber Communication Conference*, 2022.
- [2] P. Poggiolini, "The GN model of non-linear propagation in uncompensated coherent optical systems.", in *IEEE J. Lightw. Technol.*, vol. 30, no. 24, pp. 3857-3879, 2012.
- [3] C. Rottondi, et al., "Machine-learning method for quality of transmission prediction of unestablished lightpaths.", in *IEEE/OSA J. Opt. Comm. Netw.*, vol. 10, no. 2, pp. A286-A297, 2018.
- [4] S. Aladin, et al., "Quality of transmission estimation and short-term performance forecast of lightpaths.", in *IEEE J. Lightw. Technol.*, vol. 38, no. 10, pp. 2807-2814, 2020.
- [5] I. Sartzetakis, K. K. Christodoulopoulos, E. M. Varvarigos, "Accurate quality of transmission estimation with machine learning.", in *IEEE/OSA J. Opt. Comm. Netw.*, vol. 11, no. 3, pp. 140-150, 2019.
- [6] E. Seve, J. Pestic, Y. Pointurier, "Associating machine-learning and analytical models for quality of transmission estimation: combining the best of both worlds.", in *IEEE/OSA J. Opt. Comm. Netw.*, vol. 13, no. 6, C21-C30, 2021.
- [7] J. Müller, et al., "Estimating Quality of Transmission in a Live Production Network using Machine Learning.", in *Optical Fiber Communication Conference*, 2021.
- [8] A. Ferrari, et al., "Experimental Validation of an Open Source Quality of Transmission Estimator for Open Optical Networks.", in *Optical Fiber Communication Conference*, 2020.
- [9] H. J. Cho, et al., "Constellation-based identification of linear and nonlinear OSNR using machine learning: a study of link-agnostic performance.", in *OSA Optics Express*, vol. 30, no. 2, pp. 2693-2710, 2022.
- [10] I. Khan, et al., "Lightpath QoT computation in optical networks assisted by transfer learning.", in *IEEE/OSA J. Opt. Comm. Netw.*, vol. 13, no. 4, pp. B72-B82, 2021.
- [11] B. Shariati, et al., "Learning from the optical spectrum: failure detection and identification.", in *IEEE J. Lightw. Technol.*, vol. 37, no. 2, pp. 433-440, 2019.

# High Bandwidth and Ultra Low-Latency Near IR Communication Network for CMOS-compatible Integrated Photonics Chips

Javier Acevedo<sup>\*</sup>, Shahryar Sabouri<sup>†</sup>, Shiwei Shen<sup>\*</sup>, Christoph Keller<sup>‡</sup>, Joerg Hopfe<sup>‡</sup>, Stefan Reichmuth<sup>‡</sup>, Patrick Hobi<sup>‡</sup>, Marco Dietrich<sup>‡</sup>, Kambiz Jamshidi<sup>†</sup>, and Frank H. P. Fitzek<sup>\*§</sup>

<sup>\*</sup> Deutsche Telekom Chair of Communication Networks, Technische Universität Dresden

<sup>†</sup> Integrated Photonic Devices Group, Chair of Radio Frequency and Photonics Engineering

<sup>‡</sup> albis-elcon Systems Germany GmbH

<sup>§</sup> Centre for Tactile Internet with Human-in-the-Loop (CeTI)

E-mails: {firstname.lastname}@tu-dresden.de<sup>\*†§</sup> ; {firstname.lastname}@albis-elcon.com<sup>‡</sup>

**Abstract**— Photonic Integrated Circuit (PIC) have outperformed its electrical counterpart in terms of on-chip processing and power consumption by transmitting and computing optical signals at ultra-high speed, using less energy. However, fully functional optical transceivers are still rare. This work introduces the design and implementation of a tunable transceiver and an optical communication network on top of an external laser source, functioning in the near-Infrared (IR) range. The transceiver is composed of the integrated circuit of an optical chip, a hardware acceleration interface, and a customized Digital-to-Analog Converter (DAC) at the transmitter and receiver side. The photonic chip was designed and fabricated on CMOS compatible and represents an integrated Optical Phased Array (OPA), which modulates light for multi-channel beam control. The two-dimensional beam steering relies on an array of waveguide grating couplers, which ensures high accuracy and directionality during the narrow beam radiation. In the longitudinal direction, the beam is steered by a wavelength tuning mechanism, while for the lateral direction, the lobe is controlled by a network of thermo-optical phase shifters. The hardware accelerator, a Multi-processor System-on-Chip (MPSoC), computes the source data with random coefficient in Galois Fields (GF) to generate coded packets, which flow throughout the optical beams. Additionally, the accelerator is also responsible for the selection of the input voltage within the DACs to control the optical chip. To ensure successful packet transmission when the transmitter and receiver are in motion, a dynamic sliding window protocol based on Random Linear Network Coding (RLNC) was designed and implemented at the MAC layer. Extensive simulation demonstrates that our implementation results in an average packet success rate of 93.58% with reduction in average delay.

**Index Terms**—Integrated Photonics Chip, Gilbert-Elliot (GE) Channels, MPSoC, IR Communication, RLNC, Dynamic Sliding Window (DSW), VCSEL.

## I. INTRODUCTION

Although in recent years the advances in the manufacturing, integration, and miniaturization have led to considerable increments on the processing speed of the Digital Signal Processors (DSP), performance of that technology is still hindered by

the DAC. In contrast, PIC have demonstrated to outperform its electrical counterpart in terms of on-chip processing and power consumption by transmitting and computing optical signals at ultra-high speed, using less energy. In this context, the proliferation of optical hardware technology, with wavelengths ranging from IR to Ultraviolet (UV), has been the milestone for the development of new Wireless Communication Systems (WCS). Light provide unneglectable benefits in terms of throughput and low latency, reaching bit-rates up to several Gbit/s and delays at the microseconds scale over long distances. As a result, the adoption of photonic devices is taking place, enabling new use cases such as vehicle-to-vehicle communication and the Industry 4.0 (I4.0).

Applications such as Local Access Network (LAN), Storage Area Networks (SAN), and Data Center Interconnect (DCI) are based on high speed data transfer technologies such as Gigabit Ethernet and Fibre Channel, employing optical fiber transceivers [1]. Those optical devices rely on Vertical Cavity Surface Emitting Lasers (VCSEL) and Electro-Optical Transceiver (EOT) light source technologies to guarantee high reliability and bit rates, over long distances [2]. Figure 1a depicts how optical transceivers are employed to interconnect Top-of-Racks (ToR) and servers within the backhaul of the 5th Generation Cellular Networks (5G). Besides the optical wired transceivers in servers rooms, PICs have also applicability in multi-Free Space Optical (FSO) communication. FSO data links studies suggest 850 nm VCSEL arrays [3, 4, 5] and 1550 nm OPAs [6, 7, 8]. Figure 1b illustrates how data communication links based on OPA are compelling candidates for implementing fully wireless interconnects in many industrial use-cases. Optical wireless transceivers stand out for being more secure, while enabling spatial switching in comparison to legacy communication systems.

The usage of OPAs and VCSEL arrays has received great attention in both the industry and the academia [6, 7, 9]. Following this reasoning, this contribution introduces the design and implementation of an OPA, fabricated on CMOS compatible PIC technology. By exploiting the results of previous scientific

Corresponding authors: J. Acevedo, javier.acevedo@tu-dresden.de, and S. Sabouri, shahryar.sabouri@tu-dresden.de

Manuscript received March 28, 2022; revised April XX, 2022.

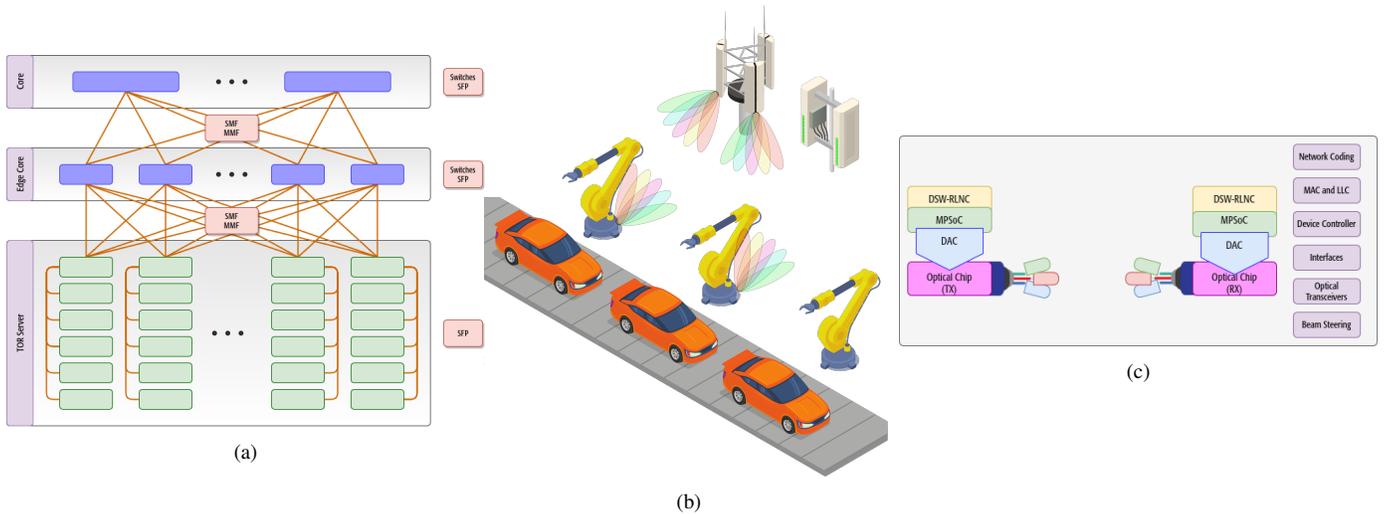


Figure 1: Figure 1a Architecture of an optical data center interconnects. Figure 1b Schematic of optical link for industrial use case. Figure 1c Schematic of the proposed communication transceiver.

works [10, 11, 12], we designed an OPA over Silicon Nitride PIC with a 850 nm wavelength, capable of performing wide beam steering of  $\pm 30^\circ$  [13]. The results of this implementation about free space data transmission were introduced in [14]. Furthermore, we compared CMOS compatible thermo-optical phase shifters in [15] to elaborate the phase shifters network of our optical chip, which is responsible for lateral beam steering of the OPA. Optimizing the performance of a single element of the phase shifters reduces the total electrical power consumption, while augmenting the speed of beam steering.

In this work, we propose a complete optical transceiver, designed to operate in the C-band (1530 – 1565 nm). Figure 1c illustrates our setup, composed by the hardware of the transceiver includes an OPA fabricated over an integrated CMOS chip; a System-on-Chip hardware module, capable of accelerating the computation of a coding scheme algorithm; and a customized DAC, which enables precise beam steering throughout thermos-optical phase shifters. In order to tackle losses and reduce the decoding delay of the of data source packets, a DSU-RLNC was designed and implemented at the MAC layer. Extensive simulation results showed that this coding scheme can match the performance of fixed window size algorithms, reaching 93.58 % of packet success rate, while reducing the average decoding delay time.

The remainder of this work is structured as follows: In section II, we provide an general view about optical chip design, coding schemes, and hardware-based signal processing accelerators used in PICs. In section III, we present in detail the components of our setup. Later, section IV introduces the parameters that were under consideration during the *mise en scène* of each module. In section V, we provide the numeric results of the discrete simulation of the DSU-RLNC. Finally, in section VI we discuss the further steps in the development of this work. In section VII, we conclude our paper.

## II. STATE-OF-THE-ART

In the literature, the design of photonic solid-state chips with hardware accelerated signal processing and coding modules have been tackled from different independent approaches. Nevertheless, it can be tedious to find scientific contributions that comprehend the complete design and deployment chain as a whole.

In the field of the PIC with chip-scale 2D beam steering, Poulton et al. [6, 7] introduce the design of an OPA, employed in Light Detection And Ranging (LiDAR) and free-space data communication applications. Similar to our work, this setup counts with control electronics, employing DACs and Field Programmable Gate Array (FPGA)s for beam control. Although the characterization of some important metrics such as maximum ranging distance and resolution are still pending, those contributions highlighted the challenges that laser-controlled OPA technology has to guarantee small-form-factor and reduced power consumption. Nevertheless, our implementation goes a step further by introducing RLNC to mitigate the effect that packet losses have in WCS.

Solid-state phase shifters for beam control have also been investigated by Sabouri et al. [10], Sabouri and Jamshidi [11]. They comprehensively researched how power can be uni- and multi-directionally radiated through silicon-based single antennas and an array of granting couplers, enabling point-to-point communication. In this work, we employed the same considerations for the design of our optical chip, maximizing the antenna array directionality, reducing the side lobe power level, and performing the a wide beam steering via phase tuning.

One intersection between our optical chip and the latest contributions about the design of OPA is the effect of the thermal cross-talk. The works in [12, 13, 15] explore how Thermo-optical Phase Shifters (TOPS) manufactured in standard Silicon-on-Insulator (SoI) require a smaller cross-talk and

power for tuning. This principle has been investigated during the implementation of our optical chip on simulations.

Emitting lasers such as VCSEL and EOT are employed as the default technology of the optical links used in data centers. Current and future computing systems will require constantly increasing data rates, which are only achievable in optical fibers. Therefore, Tatum et al. [1], Koyama [2], Wei et al. [3] provide a deep overview of the wavelength engineering of the laser technology applied to beam steering in PIC. Those works were of our interest because our prototype also employs laser to precisely control the radiation direction of the lobes, generated within the OPA.

On the other hand, there is a plethora of scientific literature about RLNC currently available. Nevertheless, we focused our attention on coding schemes capable of tackling network dynamics in wireless lossy links. In such a manner, the work of Wunderlich et al. [16], Pandi et al. [17] shows how finite sliding window RLNC is an efficient no-feedback technique to reduce the decoding delay and computing complexity, when packet drops are present in the communication. Similarly to the research carried out for this work on simulation, the variable length sliding window RLNC has been studied by Gui et al. [18]. In this approach, instead of having a fixed window size for source packet combinations, the coded packets are generated by computing an adaptable number of data packets with random coefficients in GF.

Finally, hardware-based acceleration using FPGAs has been utilized for beam steering in some OPA designs. Specifically, in [6, 7], a beam control based on Look-up Tables (LUT), stored on RAM, has been proposed, but the authors do not disclose further details about their implementation. Although our computing system is based on a MPSoC, the work from Tao et al. [19] also has a central control systems embedded in FPGAs. Furthermore, the implementation of RLNC in specialized hardware was performed by Acevedo et al. [20], demonstrating that instruction extensions of DSP can provide high throughput and low energy consumption.

To the best of our knowledge, there is no publicly available implementation that encompass PIC, coding schemes, and signal processing throughout MPSoCs. This fact represents the motivation for the design and implementation of this work.

### III. OPTICAL TRANSCEIVER

#### A. Optical Chip

In this work, we present the design of an optical chip. The chip embeds a 1550 nm wavelength OPA, designed over an area of  $16 \times 3.5 \text{ mm}^2$  SoI. The chip was fabricated in the Advanced Micro Foundry (AMF) technology, as depicted in Figure 2a. This design stands out for having an array of 64 wave-guide grating antennas, which form an emitting aperture of  $325 \times 144 \mu\text{m}^2$ . Each waveguide grating has been optimized to provide an effective and uniform 300  $\mu\text{m}$  radiation length. The antenna spacing is  $1.45\lambda$  to increase the possible Field-of-View (FoV), while minimizing the optical cross-talk. Figure 2d depicts three sets of Grating Coupler Photodetector (GC-PD), which are implemented around the antenna array to benefit

beam alignment on the receiver side. Furthermore, from an optical input, 64 optical channels were multiplexed, using a fiber edge coupler, based on a six-stage of  $1 \times 2$  Multi-mode Interferometer (MMI) tree. The proposed OPA enables 2D beam steering. A wavelength tuning adjusts the longitudinal angle, while a thermo-optical phase tuning mechanism controls the lateral coordinates of the beam. Figure 2a presents the independent phase shifters of the channels. The shifters are distributed after the  $1 \times 2$  MMI module to guarantee a cascade phase control within each optical channel. Figure 2c shows four snake-shaped top-clad heaters to provide a semi trapezoid heat flow control along the optical channels.

The aforementioned chip is compliant with a Full-width Half Maximum (FWHM) optical beam of less than  $0.5^\circ$ . The beam steering can reach up to  $\pm 45^\circ$  at a speed of  $0.26^\circ/\mu\text{s}$

#### B. RLNC

WCS can be highly unreliable, suffering from packet losses and delay by various amounts of time. In order to tackle those challenges when data is transmitted between the sender and the receiver, it is necessary to find alternatives to encode and decode messages travelling through lossy wireless links. One of those methods, which promises improve reliability, is RLNC [21]. This scheme stands out for using and distributing randomly generated coding coefficients among the end nodes. The input original data is fractionated into  $G$ ,  $G > 1$ , source symbols, which are then grouped in generations. At the transmitter, those  $G$  source symbols are coded by the encoder, using random coefficients based on linear finite arithmetic in the GF,  $GF(2^p)$ . In such a manner, the coded symbols are linear combinations of all the source symbols within a generation. The coded symbols and the random coefficients, also denominated as *coded packets*, are then transmitted at the sender, and decoded at the receiver [17].

The size of the GF determines the linear dependency among the coded packets. Higher fields, for instance  $GF(2^8)$ , which mainly represent data structures of current computing systems, reduce the reliance at the expense of computing overhead. On the other hand, although by sending more coded packets, the dependency can be also minimized, the consequence of this operation is a significant latency increment. Therefore, it is necessary to find a trade-off that both allows decoding at the receiver even when coded packets are lost and reduces or delimits delay to fulfill hard latency requirements of the communication.

Beside its appealing benefits in terms of reliability such as the node independence for coded packet recombination, RLNC is a computing expensive task, based on matrix multiplication and inversion. Therefore, most of its applicability has been limited to specialized hardware such as DSP and MPSoC, which can exploit vectorization to ensure data parallelism, and hence, high throughput and low latency.

#### C. DSW-RLNC

Dynamic systems require dynamic resilient networks to transfer data through wireless lossy links. Therefore, with

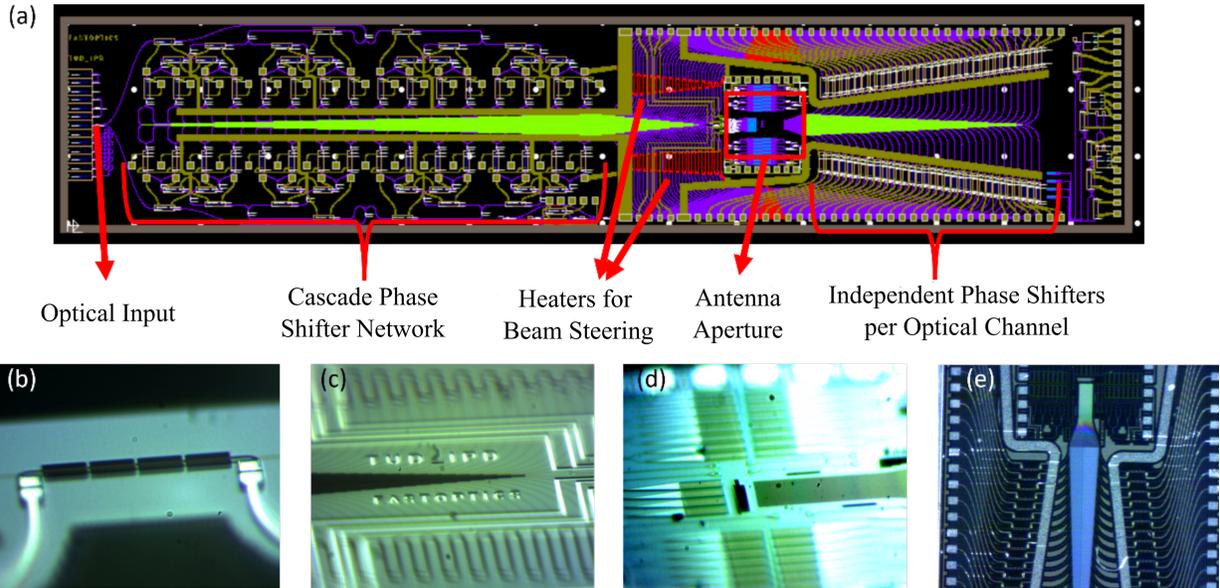


Figure 2: Figure 2a Layout of the designed and fabricated optical chip on SoI substrate, including OPA antenna and three control stages for beam steering: cascade phase shifter network, trapezoid heaters, and independent phase shifters per optical channel. Figure 2b presents single thermo-optical phase shifter element. Figure 2c depicts the trapezoid heaters along with the wave-guide routing. Figure 2d shows the emitting area. Finally, Figure 2e. illustrates the independent phase shifters for each optical channel.

transmitter and receiver under motion, the RLNC coding scheme necessitates of an adaptation to tackle the dynamic of the system. Many laser-based optical uni-cast communication applications are based on unidirectional signal propagation [22]. Therefore, when the position either of the transmitter or the receiver varies, the error rate increments considerably, making necessary packet re-transmissions. The transmitter must wait for an acknowledgment from the receiver to start the sending of the lost packets, causing the overall latency to increment [23]. In this context, we design a DSW-RLNC protocol based on the Caterpillar Random Linear Network Coding (CRLNC) [16], which has been implemented on the MAC Layer. CRLNC was designed with a fixed window size, which is intended for channels with a stable error rate. In dynamic systems, the error rate presents significant variations with respect to the time, making the fixed window size approach incapable of providing high throughput and low latency during the packet decoding. In contrast, DSW-RLNC can deal with the requirements of Mobile Ad-hoc Network (MANET)s by varying the number of source packets used to create coded packets within the encoder. Table I depicts the format of coded packets under DSW-RLNC, which consists of header and payload. The source packet represents the upper layer, while the coded packet is the packet to be transmitted. A coded packet is formed by computing several source packets with random coefficients, generated by finite field arithmetic over  $GF(2^8)$ . For instance, the coded packet 11 contains four source packets, the window size is 4, and the latest source packet ID is 7. Therefore, the coded packet 11 is a linear combination of source packet 4, 5, 6 and 7 over  $GF(2^8)$ .

TABLE I: Coded packet format for DSW-RLNC.

Pkt. ID	Header		Payload
	Last Source Pkt. ID	# Pkts.	
6	4	1	4
7	5	1	5
8	5	2	$4\oplus 5$
9	6	1	6
10	7	1	7
11	7	4	$4\oplus 5\oplus 6\oplus 7$

#### D. DAC Design

Our prototype uses a 16-channel DAC, the AD5679R (16-bit resolution) from Analog Devices [24], for optical beam control. For each channel and depending on the digital level at its input, in our configuration the DAC provides a voltage range between 0 – 2.5 V to activate 127 heaters, allowing the beam selection within the optical chip.

From the manufacturer's datasheet, we observe that when a voltage of 3.3 V is applied to the input of the DAC chip, it delivers 2.5 V output voltage with a current of 22 mA, considering an operating temperature of 25 °C. Nevertheless, as in the worst working case, the temperature can be higher than 25 °C and the input voltage coming from the evaluation board can be lower than 3.3 V, so we decided to reduce the number of heaters per DAC channel. Considering that high current provides less accuracy, we opted for a design that employs all available channels inside a DAC chip. The heaters are grouped into 7 channels. A heater channel has between 1 and 64 heaters. Therefore, in our prototype we are utilizing

TABLE II: DAC, Channels and Heaters configuration.

Channel	Heaters	DACs	Heater/DAC	Max. Current[mA]
1	1	1	1	2
2	2	1	2	4
3	4	1	4	8
4	8	2	4	8
5	16	4	4	8
6	32	7	4×5,3×4	8 or 10
7	64	16	16×4	8

TABLE III: Hardware and Software setup for transmitting data to the optical chip.

Component	Description
CPU	Intel Core i7 6700
RAM	2x16 GB
OS	Ubuntu 20.04.4LTS (GNU/Linux 5.13.0-37)
MPSoC	Xilinx ZCU102 - SFP+ Cages
Software	Xilinx Vivado and Vitis 2019.2

two DAC chips to support a total of 127 heaters, placing 4 or 5 heaters per channel. In such a manner, it is possible to apply the same voltage at every channel within the DAC chip simultaneously. Table II presents the current arrangement between DAC channels and heaters. As it is depicted, we employed up to 5 heaters per DAC channel. This limits the output current per DAC channel to a maximum of 10 mA, which makes our design more resilient to changes in the supply voltage.

### E. Hardware-Software Co-Design

All variants of RLNC has been characterized by the usage of intensive computing resources. Therefore, for this work, we have employed the Xilinx’s ZCU102 development kit. This kit is composed by a MPSoC and several peripherals such as SFP+ cages. The MPSoC will be employed for the parallelization of the matrix multiplication and inversion operations of the DSW-RLNC algorithms, while the SFP+ Ethernet interface is utilized for data management going to and coming from the optical chip. Table III summarizes the setup employed for the data transmission between the ZCU102 development kit and the extension DAC board. In contrast to legacy FPGAs, the ZCU102 embeds a complex chip composed by Programming System (PS) and Programmable Logic (PL) [25]. The PS is mainly composed of two quad-core ARM processors, the ARM Cortex A-53 and the Cortex-R5F, for application and real-time processing, respectively. Conversely, the PL contains the acceleration fabric, where we can exploit data vectorization for the parallelization of our algorithm. The AXI high-speed communication interface establishes the data transfer between the PS and the PL.

TABLE IV: Parameters for the implementation of RLNC coding schemes.

Coding	Window Size	Code Rate	Galois Field
CRLNC-1	4	1/2	$GF(2^8)$
CRLNC-2	6	1/2	$GF(2^8)$
DSW-RLNC	64-4	1/2	$GF(2^8)$

## IV. METHODOLOGY

### A. DSW-RLNC

Through a discrete event simulator, we assess the performance of the DSW-RLNC algorithm. The simulator models the sender-receiver setup based on the GE channel [3]. The channel has two states: the *good* state representing the successful arrival of all packets and the *bad* state representing the loss of all packets. In the simulation, we set a channel’s transition probability of 1/16 from the *good* to the *bad* state. In contrast, the transition probability from the *bad* to the *good* state increases linearly from 1/16 to 1/2. In this case, the average packet error rate of the channel decreases from 50 % to 6.25 %. We run the simulation 100 times, using 105 source packets.

### B. MPSoC Networking

The MPSoC has an Ethernet module based on the Gigabit Ethernet Manager (GEM), which can be set at the PS via the Extended Multiplexed I/O (EMIO) and multiplexed I/O (MIO) interfaces. Four different GEMs are integrated on the PS and can be configured independently one from another through the Reduced Gigabit Media Independent Interface (RGMII) PHY. The PL can be also employed to route Ethernet packets through the EMIO interface. For this work, we designed a Ethernet transceiver using the peripherals of the ZCU102 based on the manufacturer documentation and implementations in [26, 27]. In such a manner, we made a 1Gbps Ethernet processing system based on the PL. Figure 3 presents the building blocks of this implementation.

## V. EVALUATION

### A. DSW-RLNC

The average source packet success rate for the different methods is depicted in Figure 4. Redundant Packet Transmission (RPT) indicates the case where the source packet is simply transmitted redundantly one more time over the communication channel. Table IV presents the key parameters of CRLNC-1, CRLNC-2 and DSW-RLNC. In general, RLNC-based coding schemes outperform RPT. Among the three RLNC-based schemes, CRLNC-1 has a lower packet success rate because of the smaller window size, which prevents CRLNC-1 from recovering the source packet in case of consecutive packet loss. CRLNC-2 has the highest packet success rate 94.07 % due to its larger window size. Our proposed DSW-RLNC is only slightly lower than CRLNC-2, with a success rate of 93.58 %.

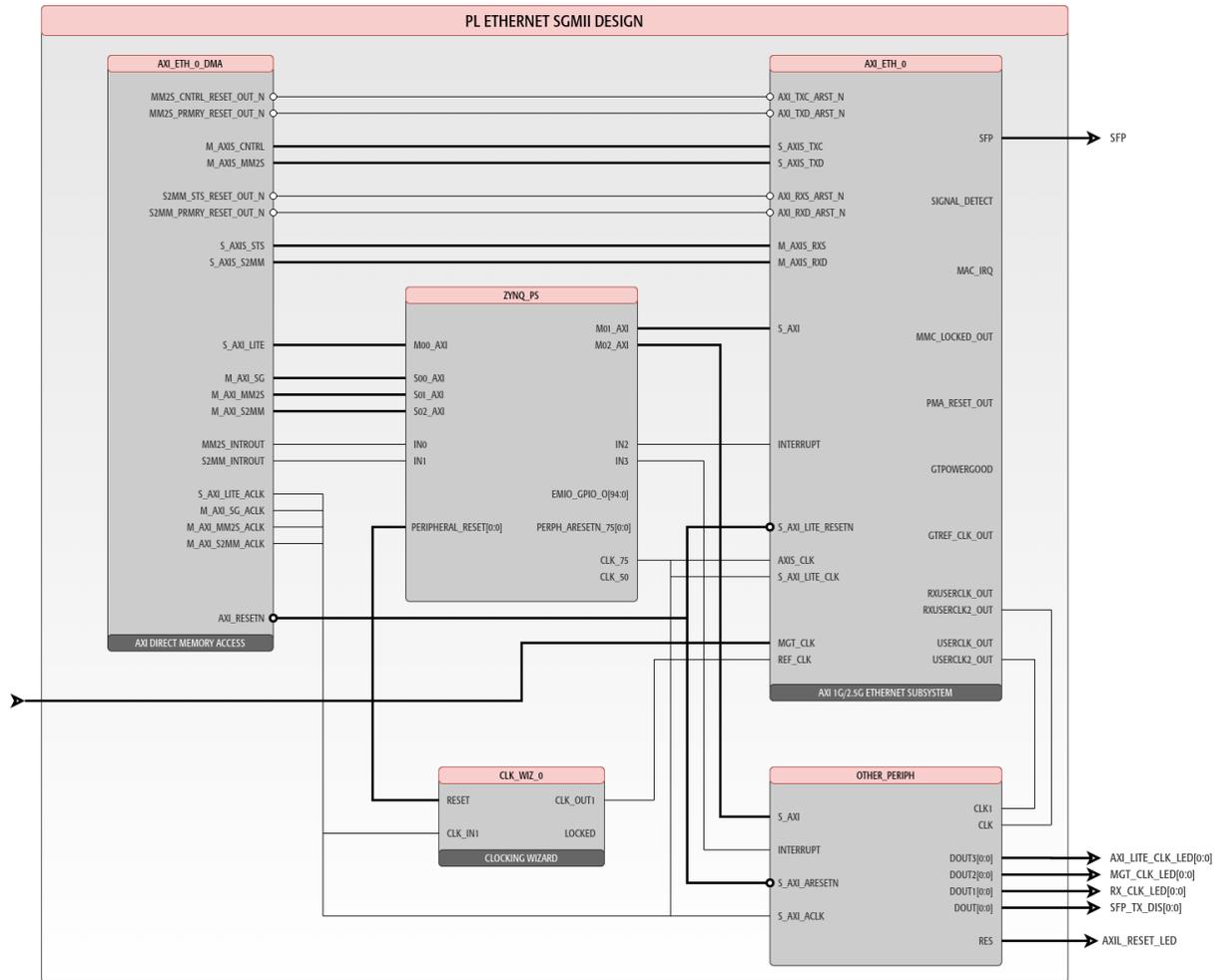


Figure 3: Building blocks of the PL-based Ethernet transceiver.

Figure 4 also illustrates the average delay of different methods. The delay of a packet is obtained by subtracting the time slot of packet arrival from the time slot of the successfully decoded source packet. Since RPT does not involve any encoding and decoding of the source packet, then the average delay of RPT is very low. Our findings indicate that among the three RLNC-based coding schemes, CRLNC-1 presents the lower average delay at the expense of lower packet success rate. Compared to CRLNC-2, DSW-RLNC has a lower average latency, while their average packet success rates are similar, around 4 to 5 time slots on average.

### B. MPSoC Ethernet

Once the design of the Ethernet transceiver finished within the IP Integrator of Xilinx's Vivado, we run the implementation and mapping to obtain metrics about the utilized processing units and memory. In such a manner, Figure 5 depicts the place and route of our design in hardware. By using Xilinx's profiling tools, we obtained that this implementation consumes 3.94 % of the LUT, 0.69 % of the LUTRAM, 2.84 % Flip-Flop (FF) and 2.47 % of the BRAM total resources.

## VI. DISCUSSION AND FUTURE WORK

There are valuable lessons learnt from the elaboration of this work. We showed that by selecting dynamic window sizes for the computing of coded packets, the decoding delay is reduced, guaranteeing on average at the same time low error rates. This paper represents an introduction and preliminary result of the implementation of a PIC, capable of transmitting computed coded packets over wireless links. Although we have presented significant progress in the design of each individual module of the transceiver, the fully integration of this setup is currently taking place. As soon as the fabrication of the Printed Circuit Board (PCB) for the chip's bring-up and wiring is completed, we will provide the characterization of the OPA, introducing metrics about power directionality and transition speed during beam steering. On the signal processing part, the incoming contributions are two-fold. Firstly, we design the first in its class DSW-RLNC encoder and decoder chip on MPSoC. By exploiting the flexibility of the PS to write high-level language code for memory control, and the high performance of the PL; we will demonstrate that this architecture fulfills the computing requirements for its implementation in industrial

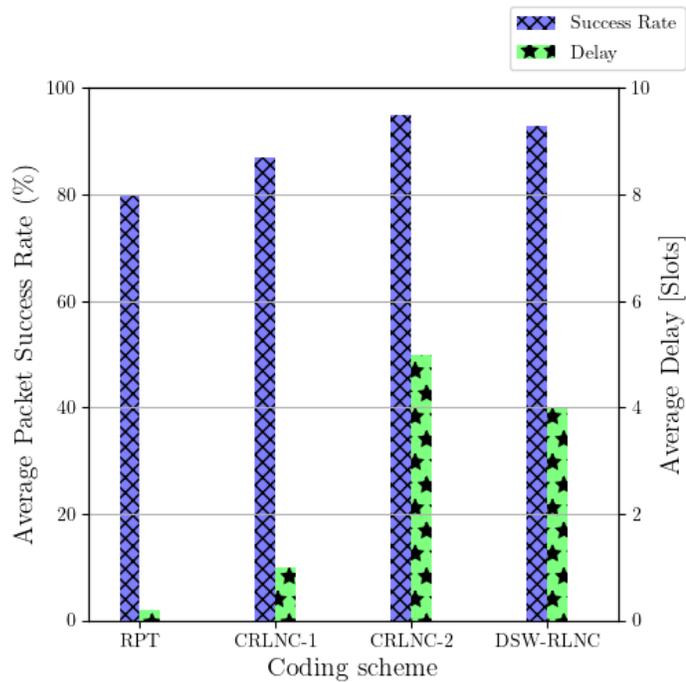


Figure 4: Simulation of the average packet success rate and average delay of the DSW-RLNC algorithm.

use-cases. Secondly, we provide the DAC characterization of the voltage switching within the OPA, depicting the angle correspondence between each channel and the lobe formed at the PIC. The source code of our implementation will be publicly available for consultation and referencing.

## VII. CONCLUSIONS

In this paper, we have presented the design and implementation of a 64-channel high-performance silicon photonic transceiver, fabricated on CMOS technology. This optical chip provides 2D beam control throughout a thermo-optical OPA, in which the transversal and vertical angles are precisely steered. The optical chip is capable of providing a beam with a FWHM of less than  $0.5^\circ$ , enabling beam steering of  $\pm 45^\circ$  at a speed of  $0.26^\circ/\mu\text{s}$ . An extension of this work will introduce them altogether with a functioning coupled setup. To improve reliability of the data packets transmitted over the FSO link, we employed RLNC. According to the simulation results of the coding scheme, we have found that DSW not only provides similar packet success rate than state-of-the-art RLNC-based methods, but also ensures lower decoding delay. From the hardware part, we observed that data transfers within the MPSoC can take place only at the PS, reducing the design complexity of the Ethernet interface. Nevertheless, if performance is under consideration, then the usage of the PL becomes compulsory.

## ACKNOWLEDGEMENT

The project underlying this publication was supported by the Federal Ministry of Education and Research of Germany

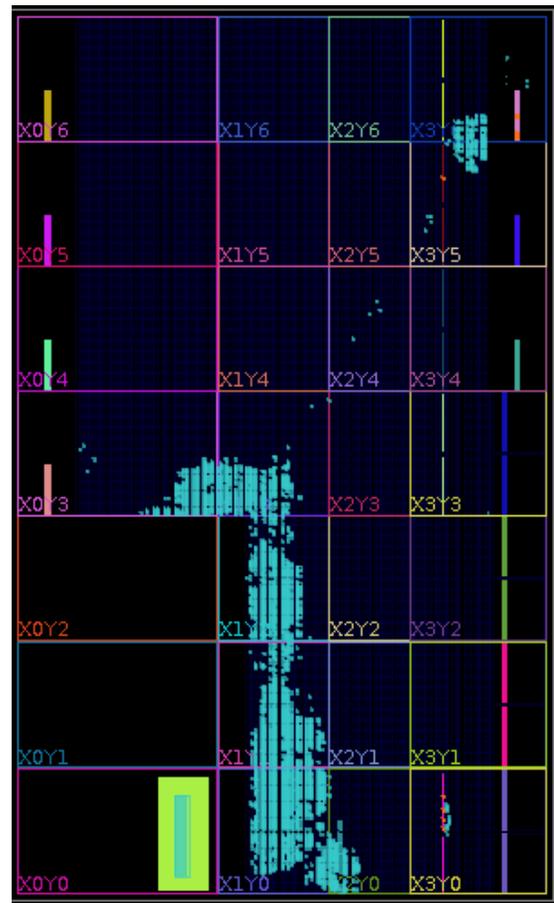


Figure 5: Place and route of the PL-based 1Gbps transceiver.

(BMBF) within the programme “Zwanzig20 – Partnership for Innovation” as part of the project consortium “fast” (funding reference number 03ZZ0532D). Additionally, this work has been also partially funded by the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft) as part of Germany’s Excellence Strategy – EXC 2050/1 – Project ID 390696704 – Cluster of Excellence “Centre for Tactile Internet with Human-in-the-Loop” (CeTI) of Technische Universität Dresden.

## REFERENCES

- [1] J. A. Tatum, D. Gazula, L. A. Graham, J. K. Guenter, R. H. Johnson, J. King, C. Kocot, G. D. Landry, I. Lyubomirsky, A. N. MacInnes, E. M. Shaw, K. Balemarchy, R. Shubochkin, D. Vaidya, M. Yan, and F. Tang, “Vcsel-based interconnects for current and future data centers,” *Journal of Lightwave Technology*, vol. 33, no. 4, pp. 727–732, 2015.
- [2] F. Koyama, “Recent advances of vcsel photonics,” *Journal of Lightwave Technology*, vol. 24, no. 12, pp. 4502–4513, 2006.
- [3] Z. Wei, M. Guan, Z. Zang, and H. Y. Fu, “Utilization of 850 nm near-infrared vcsel for high-capacity indoor free space optical communications,” in *2018 Asia Communications and Photonics Conference (ACP)*, 2018, pp. 1–3.
- [4] S. Liverman, H. Bialek, A. Natarajan, and A. X. Wang, “Vcsel array-based gigabit free-space optical femtocell communication,” *Journal of Lightwave Technology*, vol. 38, no. 7, pp. 1659–1667, 2020.
- [5] H. Kazemi, E. Sarbazi, M. D. Soltani, M. Safari, and H. Haas, “A tb/s indoor optical wireless backhaul system using vcsel arrays,” in *2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications*, 2020, pp. 1–6.

- [6] C. V. Poulton, M. J. Byrd, P. Russo, E. Timurdogan, M. Khandaker, D. Vermeulen, and M. R. Watts, "Long-range lidar and free-space data communication with high-performance optical phased arrays," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 25, no. 5, pp. 1–8, 2019.
- [7] C. V. Poulton, P. Russo, B. Moss, M. Khandaker, M. J. Byrd, J. Tran, E. Timurdogan, D. Vermeulen, and M. R. Watts, "Small-form-factor optical phased array module for technology adoption in custom applications," in *2019 Conference on Lasers and Electro-Optics (CLEO)*, 2019, pp. 1–2.
- [8] W. S. Rabinovich, P. G. Goetz, M. Pruessner, R. Mahon, M. S. Ferraro, D. Park, E. Fleet, and M. J. DePrenger, "Free space optical communication link using a silicon photonic optical phased array," in *Free-Space Laser Communication and Atmospheric Propagation XXVII*, H. Hemmati and D. M. Boroson, Eds., vol. 9354, International Society for Optics and Photonics. SPIE, 2015, pp. 96 – 101. [Online]. Available: <https://doi.org/10.1117/12.2077222>
- [9] J. A. Tatum, D. Gazula, L. A. Graham, J. K. Guenter, R. H. Johnson, J. King, C. Kocot, G. D. Landry, I. Lyubomirsky, A. N. MacInnes, E. M. Shaw, K. Balemarthy, R. Shubochkin, D. Vaidya, M. Yan, and F. Tang, "Vesel-based interconnects for current and future data centers," *Journal of Lightwave Technology*, vol. 33, no. 4, pp. 727–732, 2015.
- [10] S. Sabouri, M. Namdari, S. Hosseini, and K. Jamshidi, "1-d array of silicon nitride grating couplers for visible light communications," in *2016 IEEE International Conference on Wireless for Space and Extreme Environments (WiSEE)*, 2016, pp. 73–76.
- [11] S. Sabouri and K. Jamshidi, "Design considerations of silicon nitride optical phased array for visible light communications," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 24, no. 6, pp. 1–7, 2018.
- [12] H. Beyranvand, M. Javad Emadi, S. Sabouri, K. Jamshidi, and F. H. Fitzek, "Converged 5g and fiber-wireless access networks enhanced with visible light communications and steerable infrared beam," in *2019 2nd West Asian Colloquium on Optical Wireless Communications (WACOWC)*, 2019, pp. 152–156.
- [13] S. Sabouri, M. T. Catuneanu, L. A. M. Velasco, M. T. Fathi, and K. Jamshidi, "End-fire optical phased array with a wide beam-steering for near-visible infrared applications," in *2021 Conference on Lasers and Electro-Optics Europe European Quantum Electronics Conference (CLEO/Europe-EQEC)*, 2021, pp. 1–1.
- [14] S. Sabouri, L. A. Mendoza Velasco, and K. Jamshidi, "Near-visible end-fire array with large fov scalable for 3d sensing and fso communication," *Conference on Lasers and Electro-Optics*, 2022.
- [15] S. Sabouri, L. A. Mendoza Velasco, M. Catuneanu, M. Namdari, and K. Jamshidi, "Thermo optical phase shifter with low thermal crosstalk for soi strip waveguide," *IEEE Photonics Journal*, vol. 13, no. 2, pp. 1–12, 2021.
- [16] S. Wunderlich, F. Gabriel, S. Pandi, F. H. Fitzek, and M. Reisslein, "Caterpillar rlnc (crlnc): A practical finite sliding window rlnc approach," *IEEE Access*, vol. 5, pp. 20 183–20 197, 2017.
- [17] S. Pandi, F. Gabriel, J. A. Cabrera, S. Wunderlich, M. Reisslein, and F. H. P. Fitzek, "Pace: Redundancy engineering in rlnc for low-latency communication," *IEEE Access*, vol. 5, pp. 20 477–20 493, 2017.
- [18] C. Gui, B. Sun, Y. Song, K. Xiao, and C. Huang, "Variable length sliding window-based network coding algorithm in manets," in *Proceedings of the 14th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, ser. MobiQuitous 2017. New York, NY, USA: Association for Computing Machinery, 2017, p. 488–493. [Online]. Available: <https://doi.org/10.1145/3144457.3152354>
- [19] M. Tao, T. Peng, C. Ding, J. Guan, Y. Li, L. Zhang, B. Chena, L. Gao, X. Li, J. Song, and F. Gao, "A large-range steering optical phased array chip and high-speed controlling system," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–12, 2022.
- [20] J. Acevedo, R. Scheffel, S. Wunderlich, M. Hasler, S. Pandi, J. Cabrera, F. H. P. Fitzek, G. Fettweis, and M. Reisslein, "Hardware acceleration for rlnc: A case study based on the xtensa processor with the tensilica instruction-set extension," *Electronics*, vol. 7, no. 9, 2018. [Online]. Available: <https://www.mdpi.com/2079-9292/7/9/180>
- [21] T. Ho, M. Médard, R. Koetter, D. R. Karger, M. Effros, J. Shi, and B. Leong, "A random linear network coding approach to multicast," *IEEE Transactions on Inf. Theory*, vol. 52, no. 10, pp. 4413–4430, 2006.
- [22] X. Wang and D. Chen, "Link design for wireless optical communication network based on ant colony algorithm," *Journal of Communications and Networks*, pp. 1–8, 2022.
- [23] F. H. Fitzek, S.-C. Li, S. Speidel, T. Strufe, M. Simsek, and M. Reisslein, *Tactile internet: With human-in-the-Loop*. Academic Press, 2021.
- [24] *16-Channel, 12-/16-Bit nanoDAC+ with 2 ppm/°C Voltage Reference Temperature Coefficient, SPI Interface*, Analog Devices, 2019, rev. B. [Online]. Available: <https://www.analog.com/media/en/technical-documentation/data-sheets/AD5674-5674R-5679-5679R.pdf>
- [25] J. Acevedo. (2021) MPSoC networking and benchmarks. [Online]. Available: [https://github.com/jracevedob/MPSoC\\_Networking](https://github.com/jracevedob/MPSoC_Networking)
- [26] X. Inc. (2019) ZCU102 ps and pl based 1g/10g ethernet. [Online]. Available: <https://github.com/Xilinx-Wiki-Projects/ZCU102-Ethernet>
- [27] *PS and PL-Based 1G/10G Ethernet Solution*, Xilinx Inc., 2019, v1.5. [Online]. Available: [http://xilinx.eetrend.com/files/2019-11/wen\\_zhang\\_100046106-84570-xapp1305-ps-pl-based-ethernet-solution.pdf](http://xilinx.eetrend.com/files/2019-11/wen_zhang_100046106-84570-xapp1305-ps-pl-based-ethernet-solution.pdf)

# Distributed Sensing of Single Mode Fibers with Correlation Techniques

Florian Azendorf  
Advanced Technology  
ADVA Optical Networking SE  
Meiningen, Germany  
[fazendorf@adva.com](mailto:fazendorf@adva.com)

André Sandmann  
Advanced Technology  
ADVA Optical Networking SE  
Meiningen, Germany  
[asandmann@adva.com](mailto:asandmann@adva.com)

Michael Eiselt  
Advanced Technology  
ADVA Optical Networking SE  
Meiningen, Germany  
[meiselt@adva.com](mailto:meiselt@adva.com)

Bernhard Schmauss  
Institute of Microwaves and Photonics  
Friedrich-Alexander-Universität  
Erlangen-Nürnberg  
Erlangen, Germany  
[bernhard.schmauss@fau.de](mailto:bernhard.schmauss@fau.de)

**Abstract—** In this paper, we report on the development progress of correlation-based optical time domain reflectometry (OTDR). Substituting the direct detection receiver with a coherent receiver enables to extract the phase and polarization information of the reflected signal. Furthermore, due to the mixing of a weak probe signal with a strong local oscillator the sensitivity of the receiver improved. This improvement was demonstrated by analyzing the reflection from an angled physical contact (APC) connector. To further quantify the improvements, we compare the direct detection correlation OTDR (C-OTDR) with the coherent detection correlation OTDR (CC-OTDR) with respect to the spatial and amplitude resolution. **Keywords—**distributed sensing, OTDR, correlation techniques, phase noise,

## I. INTRODUCTION

The usage of optical fibers as a sensing medium has attracted considerable interest due to the promising properties of fibers like high sensitivity, operation in harsh environments and the immunity to electromagnetism. In recent years, also network carriers have gained interest to use their deployed fibers as combined transport and sensing media. The carrier obtains information about the surrounding fiber environment, which can be used for operational purposes or to gain further revenue. As another approach besides distributed sensing of the telecom fiber itself, fiber Bragg gratings (FBG) can be inscribed in the core for a quasi-distributed sensing. A promising technique is the use of ultra-weak FBGs (UWFBG) [1], which have a reflectivity in the order of -30 dB per grating and can thus be arranged in series to form a sensor array. The reflection of each grating leads to a discrete sampling point along the fiber, from which the reflection is significantly higher than Rayleigh backscattering. However, when thousands of gratings are cascaded, the probe signal experiences significantly higher attenuation, as compared to Rayleigh backscattering. An appropriate means to interrogate deployed fibers is OTDR. The technique measures the Rayleigh backscattering and Fresnel reflections by inserting an optical pulse into the fiber under test. Systems based on direct detection have been used for monitoring in long-haul fiber networks for more than 30 years [2]. These systems analyze the amplitude of the reflected and backscattered light. In addition to the amplitude, coherent detection enables the measurement of phase and polarization of the optical signal,

being backscattered over the fiber length. Consequently, it is feasible to obtain further information about the fiber and the environment [3]. In both cases, a rectangular shaped pulse is used to probe the fiber. Alternatively, code sequences can be used as the probe signal, improving the spatial resolution [4], [5]. In this work we will compare, the correlation-based direct detection method with the coherent method. In section II, the measurement principle and setup of the coherent correlation technique are introduced. Section III, shows the result of the fingerprint measurement, and compares both methods regarding their spatial resolution and amplitude sensitivity. In section IV, we will discuss the laser phase noise as one of the main issues impacting the measurement, when using telecommunication components. Finally, we conclude our findings and present an outlook in section V.

## II. GENERAL MEASUREMENT SETUP

The general measurement setup of the coherent correlation OTDR is illustrated in Figure 1. A tunable laser with a typical linewidth of 25 kHz generates a continuous wave signal. One part is used as the local oscillator (LO) of the coherent receiver, while the other part is fed into a Mach-Zehnder modulator (MZM). A bipolar sequence is modulated onto the carrier, and the probe signal is sent into the fiber under test (FuT) via an optical circulator. To reduce the influence of undesired strong reflections at the input and output of the FuT, angled physical connectors (APC) are used. The reflected and backscattered signals are received with a dual-polarization integrated coherent receiver and recorded with a real-time oscilloscope. In a single shot, multiple frames of the sequence are transmitted and consecutively recorded. Similar to [5], each frame consists either of a pseudo random binary sequence (PRBS) or of two Golay sequences, followed by a zero padding. The overall frame length corresponds to the round-trip time of the FuT to obtain unambiguous results. Four signals are recorded at the output of the coherent receiver, representing the in-phase and quadrature components for x- and y-polarizations. A cross-correlation between the transmitted and the received sequences is then performed, improving the signal-to-noise ratio and the spatial resolution. Here, complementary Golay sequences exhibit perfect acyclic auto-correlation properties,

if both sequences are used and their respective correlation functions are added. After cross-correlation in the testbed setup, the frames are further processed to extract amplitude, phase, and polarization of the backscattered signals.

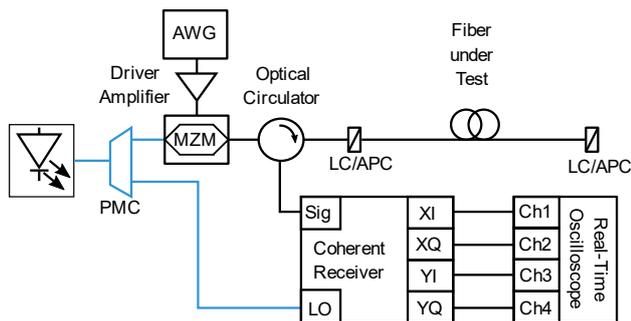


Figure 1 General schematic of the Coherent Correlation-OTDR. Blue color denotes polarization maintaining fibers.

### III. MEASUREMENTS AND COMPARISON

#### A. Fingerprint of a single mode optical fiber

Rayleigh backscattering is a result of the inhomogeneities along the fiber. These inhomogeneities originate from the amorphous character of glass. By probing the fiber with a pulse or sequence, the backscattering trace can be observed, which is a coherent combination of the backscattered fields from the individual inhomogeneities, and which is constant under steady conditions. This trace is unique for each fiber, and therefore it can be denoted as a fingerprint. The fingerprint of a fiber with length 200 m was measured with the CC-OTDR. In the experiment, we used a probe bit rate of 1.25 Gbit/s, a sampling rate of 25 GS/s, and a 511-bit PRBS. 500 consecutive frames were recorded. The recorded signals were filtered with a low pass filter and the cross correlation with the transmitted sequence was performed. To obtain the fingerprint, the absolute square of the individual correlation signals for x- and y-polarization are added. The results of the investigation are depicted in Figure 2 starting at fiber length 48.3 m.

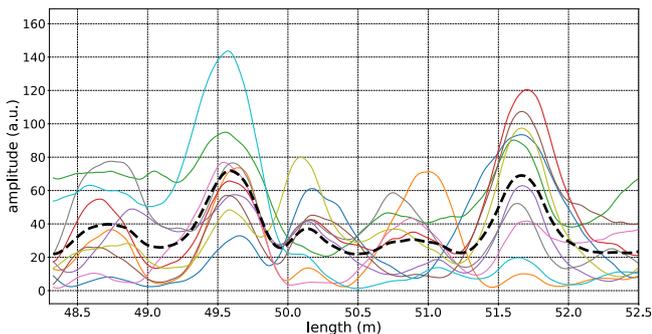


Figure 2 Fingerprint of individual frames (solid line) and averaged fingerprint of the ten frames (dashed line).

In the figure, the amplitude of each fingerprint is depicted as a function of fiber length for ten consecutive frames. Each solid line represents a recorded trace, while the dashed line shows the average trace of the ten frames. It can be seen that there is a good overlap of the peak locations of the ten frames with the average trace at 49.57 m, and 51.67 m. Additionally, weak peaks were observed at 48.715 m, 50.135 m, and

50.87 m. The variations of the amplitude are presumably caused by vibrations and noise in the laboratory.

#### B. Sensitivity

It has been highlighted in the literature that coherent detection improves the sensitivity compared to the typical OTDR with direct detection [6, 7]. The sensitivity describes the scale factor between measurand (i.e., the reflected and backscattered optical power) and the received electrical signal, and it defines, which scattering, or reflection events are resolvable. It should be noted that the term sensitivity can be misunderstood with measurement resolution. The latter is the ability of the measurement technique to distinguish small changes in the quantity that is measured [8]. The probe bit rate in the experiment was 5 Gbit/s, two 2048-bit Golay sequences were used, and the sampling rate was 25 GS/s. As fiber under test, three patch cords with different lengths were cascaded. The results of the experiment are illustrated in Figure 3 showing the averaged absolute square of the correlation signals, as a function of the fiber length.

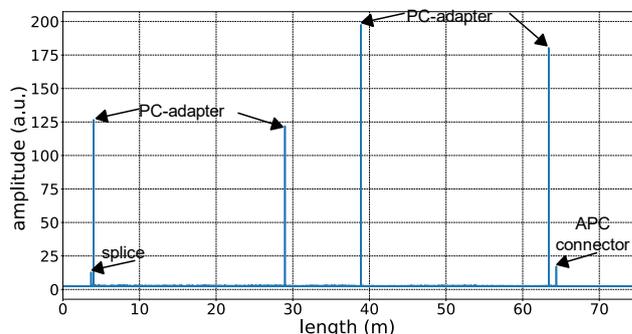


Figure 3 Averaged CC-OTDR fingerprint showing the three cascaded jumper fibers and several reflections from PC and APC connectors and a splice.

The figure shows several reflection peaks at the interface between the fibers with lengths of 25 m, 10 m, and 25 m. Furthermore, the results demonstrate the feasibility to measure APC connector reflections as well as reflections caused by splices, with both events having low reflectance in the order of -60 dB [2]. The experiment was repeated with the direct detection C-OTDR similar to our previous work [5]. The coherent receiver shown in Figure 1 was substituted by a photodiode/amplifier (PIN/TIA) combination. Since no LO is needed in the direct detection setup, the laser was directly connected to the MZM. Additionally, two Golay sequences were used to probe the fiber. The results of this experiment are illustrated in Figure 4. In the figure, the sum of the correlation functions is shown. It can be seen that the PIN/TIA combination of the direct detection method has a significantly lower sensitivity, which is noticeable since no events are visible. The reason is that the reflections of the adapters and connectors are below -46 dBm, which was measured at the optical circulator. Another reason is the vertical resolution of the real-time oscilloscope with 10 mV/div and its 8-bit quantization.

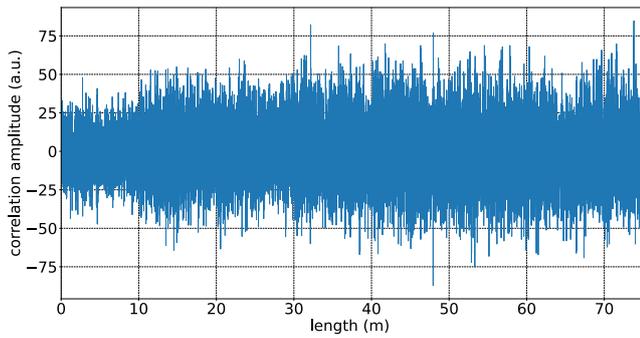


Figure 4 Amplitude of the correlation trace for direct detect C-OTDR and the same FuT as in Figure 3.

### C. Spatial Resolution

Another important parameter of OTDR measurement techniques is the spatial resolution. This parameter describes the capability of an OTDR to distinguish between reflective events with close proximity. In general, the parameter depends on the probe pulse width and the bandwidth as well as the sampling rate of the receiver structure. By considering all parameters, the spatial resolution of both techniques can be estimated and measured. The spatial resolution of the C-OTDR with direct detection was determined by using two fibers with roughly the same length as FuT coupled together with a 3 dB coupler. In the experiment the probe bit rate was varied between 10 Gbit/s, 5 Gbit/s and 2.5 Gbit/s, and the sampling rate was 50 GS/s. The overall bandwidth of the photodiode and the real-time oscilloscope was 10 GHz. Consequently, the temporal resolution should be 100 ps, 200ps and 400 ps, corresponding to a spatial resolution of 10 mm, 20 mm, and 40 mm, respectively. The results in Figure 5 show the measured trace with the different probe bit rates.

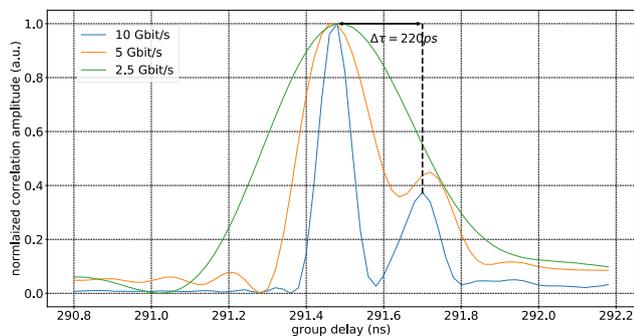


Figure 5 Spatial resolution analysis with different probe bit rates of the correlation sequence with direct detection.

From the measurements it can be seen that with 5 Gbit/s and 10 Gbit/s data rates it is feasible to measure a difference of 220 ps, which corresponds to a length of 2.2 cm between two reflection events. To investigate the resolution of the coherent correlation OTDR, a UWFBG array inscribed in a 106 m fiber was used with a grating separation of 50 mm. Each grating had a length of 10 mm, and the reflectivity was -30 dB each. The results of the experiment are illustrated in Figure 6 showing the correlation signal of the first 100 gratings with different amplitudes. The inset of the figure shows the whole fiber length with 2000 gratings, sampled at a frequency of 193.4 THz for that particular measurement. The results show the reflection for one sample frequency of

each FBG. Since the Bragg wavelength is different for each FBG the correlation amplitude differs.

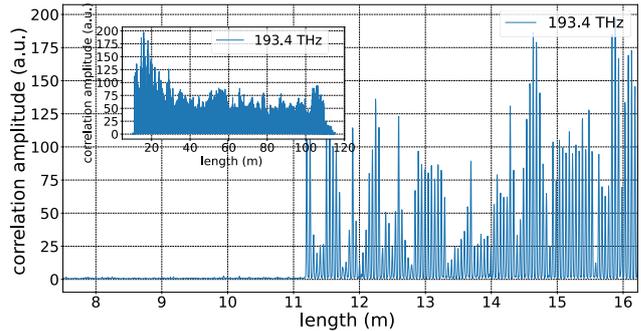


Figure 6 Correlation signal of the first 100 gratings of the FuT probed with a carrier frequency of 193.4 THz. Inset: Signal of the whole FuT with 2000 grating.

### IV. LASER PHASE NOISE INFLUENCE

In order to obtain optimum results from the correlation in the coherent detection setup, the laser phase difference between the LO and the reflected signal needs to be stable over the received sequence pattern. Hereinafter, an experiment is conducted that highlights the effect of laser phase noise on the correlation result. Two fibers under test are interrogated with the CC-OTDR setup. The first is a short fiber of 8 meter length with an open PC connector at the fiber end (i.e. transition from silica to air) to generate a strong reflection. Secondly, a 10.14 km fiber is used, which has the same open connector at the fiber end. A set of Golay sequences of length 4096 bits is transmitted with a symbol rate of 1.25 GBaud in a time-multiplexed fashion. The respective correlation functions of each complementary sequence are combined to obtain the total correlation trace. The resulting normalized total correlation magnitudes of the end-face reflections for the two fiber lengths are shown in Figure 7.

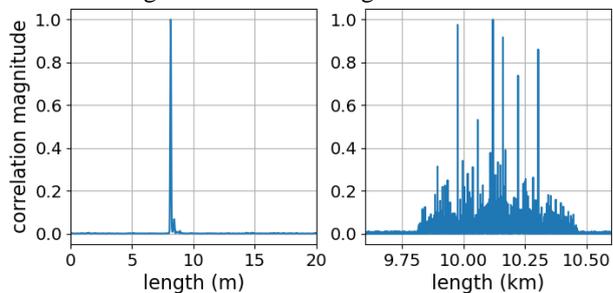


Figure 7 Normalized correlation magnitude (x-polarization only) of an open PC connector after 8 meters (left) and 10.14 km (right)

As expected for the short fiber of 8 m length, a clear correlation peak is visible. In contrast, the reflection originating from the end of the 10.14 km fiber results in a correlation pattern with multiple peaks and high side lobe noise. The latter is a result of a fast drift of the phase difference between the reflected signal and the LO over the received correlation sequence. Since a single-laser setup with self-homodyne detection is used, as shown in Figure 1, the superimposed LO and signal at the receiver are affected by different laser phase noise evolutions due to their different delays. In addition, the laser that is implemented in this experiment exhibits a slow modulation of its instantaneous frequency. Both effects combined lead to the fast drift of the phase difference when considering the reflection from 10.14 km. Considering a single code sequence, the phase difference

needs to be stable over the received sequence. Using two complementary sequences, the phase difference needs to be stable such that both sequences are measured with the same phase difference.

## V. SUMMARY AND CONCLUSION

In this work, we presented a comparison between a direct detection and a coherent correlation OTDR method. It was shown that the sensitivity of the method was improved due to coherent detection. Furthermore, the spatial resolutions were compared, and it was shown that the parameter depends on the probe bit rate and the overall bandwidth of the receiver.

## ACKNOWLEDGMENT

This work was partly funded by the German Federal Ministry of Education and Research (FKZ16KIS1279K) in the framework of the CELTIC-NEXT project AI-NET-Protect (Project ID C2019/3-4).

## REFERENCES

- [1] M. Wu, C. Li, X. Fan, C. Liao and Z. He, "Large-scaled multiplexed weak reflector array fabricated with a femtosecond laser for a fiberoptic quasi-distributed acoustic sensing system," *Optical Letters*, pp. 3685-3688, 1 Juli 2020.
- [2] D. Anderson, L. Johnson and F. Bell, *Troubleshooting Optical Fiber Networks-Understanding and Using Your Optical Time-Domain Reflectometer*, Academic Press Elsevier, 2004.
- [3] E. Ip, J. Fang, Y. Li, Q. Wang, M. F. Huang, M. Salemi and Y. K. Huang, "Distributed Fiber Sensor Network Using Telecom Cables as Sensing Media Technology Advancements and Applications," *Journal of Optical Communications and Networking*, vol. 14, no. 1, pp. A61-A68, 2022.
- [4] C. Dorize and E. Awwad, "Enhancing the Performance of Coherent OTDR Systems with Polarization Diversity Complementary Codes," *Optics Express*, vol. 26, pp. 12878-12890, 2018.
- [5] F. Azendorf, A. Dochhan and M. H. Eiselt, "Accurate Single-Ended Measurement of Propagation Delay in Fiber Using Correlation Optical Time Domain Reflectometry," *Journal of Lightwave Technology*, pp. 5744-5752, 15 September 2021.
- [6] A. H. Hartog, *An Introduction to Distributed Optical Fibre Sensors*, Boca Raton: CRC Press, 2017.
- [7] L. S. Grattan and B. T. Meggitt, *Optical Fiber Sensor Technology: Advanced Applications-Bragg Gratings and Distributed Sensors*, Boston, MA: Springer, 200.
- [8] JCGM, "International Vocabulary of Metrology–Basic and General Concepts and Associated Terms," in *Chemistry International*, 2012.

# ITG-Fachberichte

Die Beiträge der ITG-Fachtagungen und -Workshops sind in den ITG-Fachberichten dokumentiert. Die von Programmausschüssen selektierten und zumeist reviewten Beiträge stellen aktuelle Entwicklungen aus der Informationstechnik vor.

VDE

VERLAG

Technik. Wissen.  
Weiterwissen.



ITG-Fachbericht Band 304

## Mobilkommunikation Technologien und Anwendungen

Vorträge der 26. ITG-Fachtagung, 18. – 19. Mai 2022 in Osnabrück

► Das Schwerpunktthema der diesjährigen Tagung lautet „5G-Lösungen und 6G-Ausblick“. Dieses Thema ist auf der Tagung in vier Key Notes adressiert und in sieben Sitzungen diskutiert worden. Die Sitzungen widmen sich den Themen 6G, 5G-Evolution, Campusnetze und Industrie 4.0, Low Power Wide Area Networks, Small Cells und DECT NR, resiliente Netze sowie der Funkversorgung in kritischen Regionen.

2022, 130 Seiten

79,- €

ISBN 978-3-8007-5873-9 [Buch] ISBN 978-3-8007-5874-6 [E-Book]

Titel / Untertitel	Jahr	Seiten	Buch/CD/E-Book	ISBN-Nummer	Preis
<b>ITG-Fachbericht Band 303 – Sensoren und Messsysteme</b> Beiträge der 21. ITG/GMA-Fachtagung, 10. – 11. Mai 2022 in Nürnberg	2022	534	- / + / +	978-3-8007-5835-7 (CD-ROM) 978-3-8007-5836-4 (E-Book)	ca. 230,-
<b>ITG-Fachbericht Band 302 – MBMV 2021</b> Methoden und Beschreibungssprachen zur Modellierung und Verifikation von Schaltungen und Systemen 25. Workshop, 18. – 19. Februar 2022, Hochschule Niederrhein, Online	2022	72	- / + / +	978-3-8007-5754-1 (CD-ROM) 978-3-8007-5755-8 (E-Book)	48,-
<b>ITG-Fachbericht Band 301 – Kommunikationskabelnetze</b> Vorträge der 28. ITG-Fachtagung, 14. – 15. Dezember 2021 in Köln	2021	151	- / + / +	978-3-8007-5709-1 (CD-ROM) 978-3-8007-5710-7 (E-Book)	87,-
<b>ITG-Fachbericht Band 300 – WSA 2021</b> 25th International ITG Workshop on Smart Antennas 10 – 12 November 2021, EURECOM, French Riviera	2021	471	- / + / +	978-3-8007-5686-5 (CD-ROM) 978-3-8007-5688-9 (E-Book)	198,-
<b>ITG-Fachbericht Band 299 – Mobilkommunikation – Technologien und Anwendungen</b> Vorträge der 25. ITG-Fachtagung, 3. – 4. November 2021 in Osnabrück	2021	139	+ / - / +	978-3-8007-5674-2 [Buch] 978-3-8007-5675-9 [E-Book]	80,-
<b>ITG-Fachbericht Band 298 – Speech Communication</b> 14th ITG Conference, 29.09. – 01.10.2021, Online-Event	2021	223	- / + / +	978-3-8007-5627-8 (CD-ROM) 978-3-8007-5628-5 (E-Book)	116,-
<b>ITG-Fachbericht Band 297 – Photonische Netze</b> Beiträge der 22. ITG-Fachtagung, 19. – 20. Mai 2021, Online-Veranstaltung	2021	91	- / + / +	978-3-8007-5555-4 (CD-ROM) 978-3-8007-5556-1 (E-Book)	52,-
<b>ITG-Fachbericht Band 296 – MBMV 2021</b> Methoden und Beschreibungssprachen zur Modellierung und Verifikation von Schaltungen und Systemen 24. Workshop, 18. – 19. März 2021, TU München, Online-Veranstaltung	2021	136	- / + / +	978-3-8007-5500-4 (CD-ROM) 978-3-8007-5501-1 (E-Book)	78,-
<b>ITG-Fachbericht Band 295 – Breitbandversorgung in Deutschland</b> Beiträge der 15. ITG-Fachkonferenz, 2. – 3. März 2021, Fraunhofer Heinrich-Hertz-Institut, Berlin	2021	55	- / - / +	978-3-8007-5475-5 (E-Book)	32,-
<b>ITG-Fachbericht Band 294 – Photonische Netze</b> Beiträge der 21. ITG-Fachtagung, 24. – 25. November 2020, Online-Veranstaltung	2020	129	- / + / +	978-3-8007-5423-6 (CD-ROM) 978-3-8007-5424-3 (E-Book)	74,-
<b>ITG-Fachbericht Band 293 – ANALOG 2020</b> Analoge Schaltungen: Schlüsselssysteme für Automotive, IoT und zukünftige drahtlose Technologien Beiträge der 17. ITG/GMM-Fachtagung, 28. – 30. September 2020, Online-Veranstaltung	2020	88	- / - / +	978-3-8007-5336-9 (E-Book)	51,-

Sämtliche Veröffentlichungen der ITG-Fachberichte können auch im Abonnement bezogen werden. Preisänderungen und Irrtümer vorbehalten.

Bestellen Sie jetzt: (030) 34 80 01-222 oder [www.vde-verlag.de/itg-fachberichte](http://www.vde-verlag.de/itg-fachberichte)



## Tagungsbände online

Erwerb von einzelnen Tagungsbeiträgen im PDF-Format

Der VDE VERLAG publiziert jährlich rund 30 Tagungsbände mit mehr als 1.500 Fachbeiträgen. Diese können Sie auch einzeln erwerben. Zurzeit stehen Ihnen mehr als 24.000 Fachbeiträge im PDF-Format zum Download zur Verfügung. Die Kosten pro Beitrag betragen 15,- €.

Recherchieren Sie auf [www.vde-verlag.de](http://www.vde-verlag.de) in einer täglich wachsenden Zahl von technisch-wissenschaftlichen Fachbeiträgen und erwerben Sie diese im direkten Download mit bequemen Bezahlmöglichkeiten per Kreditkarte oder PayPal.

Weitere Informationen: [www.vde-verlag.de/buecher/proceedings/beitraege-suchen.html](http://www.vde-verlag.de/buecher/proceedings/beitraege-suchen.html)

## Ihre Meinung zählt!

Sagen Sie uns Ihre Meinung zum aktuellen E-Book und teilen uns Ihre weiteren Informationswünsche mit. Alle Einsender nehmen an der quartalsweisen Verlosung einer LED-Taschenleuchte oder einer original VDE-Umhängetasche teil.

Einfach auf **Feedback** klicken (ggf. Online-Verbindung prüfen) und los geht's!

**VDE**

VERLAG

Technik. Wissen.  
Weiterwissen.

[www.vde-verlag.de/newsletter](http://www.vde-verlag.de/newsletter)

## Nutzungsbedingungen

Die über die Webseite des VDE-Verlages zur Verfügung gestellten elektronischen Produkte, zum Beispiel E-Books, Tagungsbeiträge, Fachpublikationen und Normen, sind urheberrechtlich geschützt. Der Kunde erhält für einen über die Webseite des Verlages erworbenes und herunter geladenes elektronisches Produkt ein einfaches, nicht übertragbares Nutzungsrecht, um das elektronische Produkt auf einem Computer (Einzelplatz) bestimmungsgemäß zu nutzen. Der Kunde ist berechtigt, von jedem elektronischen Produkt einen Ausdruck zu machen.

Darüber hinaus ist der Kunde nicht berechtigt, ein elektronisches Produkt oder Teile davon zu kopieren, zu verändern, zu bearbeiten oder über das Internet/Intranet oder irgendeiner anderen Weise Dritten zur Verfügung zu stellen. Zwingend geltendes Urheberrecht wird durch diese Ziffer nicht beschränkt.

## Terms of Use

The electronic products that are made available over the website of the Publisher are copyright protected. For an electronic product, like Conference articles, E-Books and Standards, that has been purchased and downloaded through the website of the Publisher, Publisher grants Customer a non-exclusive and nontransferable right to use such electronic product on one computer (single workstation) in compliance with the purpose of the Agreement. Customer is entitled to make one printout per electronic product.

Moreover, the customer is not entitled to copy, modify, edit the electronic product or parts thereof or make it available to third parties over the internet or otherwise. Mandatory copyright law shall not be limited by this Section.