



Università telematica delle
Camere di Commercio Italiane

**PHD PROGRAMME IN
BIG DATA AND ARTIFICIAL INTELLIGENCE**

Curriculum "Artificial Intelligence for Industry 5.0 and circular economy"

38th Cycle

*PhD Dissertation in
A Digital Surrogate Framework for End-to-End Optimization of
IM/DD Optical Communication Systems*

Dr Ramin Solaimani

Programme Coordinator
Prof. BarbaraMarini

Supervisor
Prof. Sergio Maria Patella

Co-Supervisor
Prof. /Dott. Dario Ferrillo

Academic Year 2024 / 2025



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Università telematica delle
Camere di Commercio Italiane

Acknowledgements

First and foremost, I am deeply grateful to my mentor, Prof. Luca Potì, for his guidance, insight, and the knowledge he has shared with me throughout this work. I want to express my sincere appreciation to Prof. Enio Cascetta, who served as my first tutor at the beginning of my PhD journey, for his support and guidance during that initial phase. I am also grateful to my tutor and co-tutor, Prof. Sergio Maria Patella, for his continued support and valuable advice.

I would like to acknowledge Dr. Dario Ferrillo and Dr. Agata Quattrone from Al maviva for their support, collaboration, and the effort they dedicated to my research activities.

I am also very thankful to Prof. Nicola Calabretta at Eindhoven University of Technology, ECO Group, for hosting me and for his support and wisdom. In addition, I warmly thank Ryosuke Matsumoto and Shiyi Xia for their assistance during my visit to TU/e.

I would also like to thank my colleagues and friends at Universitas Mercatorum, the TeCIP Institute of Scuola Superiore Sant’Anna, CNIT PNTLab in Pisa, and Eindhoven University of Technology for their collaboration, encouragement, and for creating a supportive and stimulating environment.

Finally, I extend my heartfelt appreciation to my wife, Pantea, for her unwavering love, patience, and support throughout this journey. I am also deeply grateful to my parents for everything they have done for me and for their constant encouragement.

Table of Contents

Chapter 1: Introduction	1
1.1 Background and motivation	1
1.2 Problem statement	2
1.3 Thesis hypothesis and objectives	3
1.4 Research questions	4
1.5 Methodological approach	5
1.5.1 End-to-end modular architecture	5
1.5.2 Device-aware impairment modeling	6
1.5.3 ML readiness and validation strategy	6
1.6 Main contributions	7
1.7 Thesis organization	8
1.8 Notation and scope of applicability	9
 Chapter 2: Industry 5.0 and Smart Urban Mobility: Circular Economy Perspectives and the Role of Optical Communication	10
2.1 Purpose and Position in the Thesis	10
2.1.1 Motivation and scope of the chapter	10
2.1.2 Relation to the thesis and doctoral programme	11
2.1.3 Organization of the chapter	12
2.2 Industry 5.0 and Circular Economy	13
2.2.1 From Industry 4.0 to Industry 5.0	13
2.2.2 Circular economy principles and policy background	15
2.2.3 How Industry 5.0 and the circular economy reinforce each other	17
2.3 Smart Cities and Smart Urban Mobility	18
2.3.1 Definitions and dimensions of a smart city	18
2.3.2 Urban traffic management as a smart-city "nervous system"	20
2.3.3 Data, AI, and digital infrastructure requirements in smart mobility	22
2.3.4 Rome as a metropolitan example and the role of industrial actors	24
2.4 Optical Communication as Enabler	26
2.4.1 ICT backbone: from sensors to data centers	26
2.4.2 Short-reach IM/DD optical links in data centers and edge computing	27
2.4.3 Energy efficiency, resilience, and circular-economy implications	28
2.5 The digital surrogate in this ecosystem	29
2.5.1 Role of the digital surrogate for an IM-DD BPAM link	29
2.5.2 Optimized BPAM links as enablers of sustainable smart-city infrastructure	30
2.5.3 Indirect link to smart-city mobility and traffic management	31
 Chapter 3: Background & State of the Art	32
3.1 IM-DD optical links: architecture and performance metrics	32
3.1.1 System architecture of short-reach IM-DD optical links	32
3.1.2 Modulation formats and performance metrics in IM-DD links	35
3.2 Device non-idealities and impairments	39
3.2.1 DAC non-idealities: quantization, linearity, dynamic errors, and jitter	39
3.2.1.1 Quantization	40
3.2.1.2 Static linearity: INL and DNL	41
3.2.1.3 Dynamic errors and glitches	43
3.2.1.4 Thermal and reference noise	44
3.2.1.5 Timing jitter	45
3.2.1.6 Implications for the digital surrogate	46
3.2.2 Semiconductor optical amplifier (SOA): gain saturation, pattern effects, ASE ...	47
3.2.2.1 SOA roles in IM/DD systems: booster vs pre-amplifier	47
3.2.2.2 Small-signal gain and gain saturation (static compression)	49

3.2.2.3 Gain dynamics and pattern effects (memory and data dependence).....	50
3.2.2.4 ASE noise and OSNR penalties in direct detection	52
3.2.2.5 Nonlinearities most relevant to IM/DD PAM and BPAM.....	54
3.2.2.6 O-band vs C-band use, and a short comparison with EDFA.....	56
3.2.3 Photodiode / TIA and ADC (bandwidth, noise, sampling, quantization).....	58
3.2.3.1 Photodetection and electrical front-end: roles and interactions.....	58
3.2.3.2 Noise mechanisms and sensitivity limits	60
3.2.3.3 Bandwidth limitations, ISI, and front-end linearity	61
3.2.3.4 ADC sampling rate, quantization, and ENOB constraints.....	62
3.2.3.5 Sampling-time errors: clock jitter and aperture uncertainty	64
3.2.3.6 Implications for a digital surrogate and ML-ready metadata	65
3.3 Existing modeling and simulation approaches	66
3.3.1 Commercial and academic link simulators, analytical models, digital twins	66
3.3.1.1 Analytical and semi-analytical link models (fast screening).....	66
3.3.1.2 System-level time-domain simulators (waveform-accurate link models) .	68
3.3.1.3 Commercial simulators and vendor-tool workflows (component co-design)	70
3.3.1.4 Digital twins in optical networks	72
3.3.2 ML/AI in optical communications	74
3.3.3 Gap analysis and positioning of this thesis.....	77
Chapter 4: Architecture of the Digital Surrogate Framework	80
4.1 Design principles	80
4.1.1 Modularity and strict separation into blocks (Tx/channel/Rx).....	80
4.1.2 Full parameterization and "no hard-coding" policy.....	81
4.1.3 Domain-aware modeling (digital/electrical/optical) and interface contracts	82
4.1.4 Calibration readiness and hardware-aligned abstraction level	83
4.1.5 ML-readiness: structured metadata and repeatable dataset generation	84
4.2 Global configuration layer	85
4.2.1 The parameters structure: hierarchy and naming conventions	85
4.2.2 Enabling/disabling blocks and features	87
4.2.3 Reproducibility controls (random seeds, scenario IDs, sweep management).....	88
4.2.4 Calibration parameters vs nominal parameters (measured-input strategy)	89
4.2.4.1 Classical parameter fitting and identifiability	90
4.2.4.2 Bayesian calibration and uncertainty-aware fitting	91
4.2.4.3 Bayesian optimization as a practical calibration engine	92
4.2.4.4 ML-driven calibration and closed-loop tuning	93
4.3 Global runtime data structures.....	94
4.3.1 The signal structure	94
4.3.2 The time structure and the time manager concept.....	96
4.3.3 Visualization metadata (visMeta) and its separation from the signal path.....	97
4.3.4 ML/optimization metadata: what is stored, where it is stored, why	98
4.3.5 Logging and diagnostics conventions	99
4.4 Block interface conventions and execution model	100
4.4.1 End-to-end chain overview.....	100
4.4.2 Standard block I/O contract	101
4.4.3 Unit conventions and domain transitions.....	102
4.4.4 How blocks update signal, time, and optional metadata.....	103
4.4.5 Lightweight execution modes	104
4.5 Visualization and "virtual instruments"	105
4.5.1 Rationale: virtual instrumentation as part of surrogate credibility.....	105
4.5.2 Time-domain visualization and summary measurements	106
4.5.3 Eye-diagram generation and quality metrics	107
4.5.4 Spectrum and constellation views	108
4.5.5 BER/performance tracking across parameter sweeps	109
Chapter 5: Transmitter-Side Modeling: Digital, Electrical, Optical	110

5.1 Chapter scope, assumptions, and links to the framework	110
5.1.1 Position in the end-to-end chain	110
5.1.2 Domain transitions and unit conventions used in this chapter	111
5.2 Digital transmit processing and symbol formation	112
5.2.1 Bit sources and reproducibility	112
5.2.2 Mapping for PAM and BiPAM/BPAM	113
5.2.3 Optional FEC placeholder and framing hooks	114
5.2.4 Pulse shaping / upsampling strategy and reference rate definition	115
5.3 DAC surrogate block (digital front-end + converter core)	116
5.3.1 Role of the DAC in the surrogate and why device-level modeling matters	116
5.3.2 Sampling-rate matching and digital conditioning inside the DAC block	118
5.3.3 Electrical-range mapping and polarity handling	119
5.3.4 Static transfer function: quantization + code mapping + clipping	120
5.3.5 Static nonlinearities beyond quantization: INL/DNL	122
5.3.6 Dynamic reconstruction behavior: ZOH + output network	124
5.3.7 Noise model set and parameterization strategy	126
5.3.8 Metadata and calibration hooks for ML/optimization	128
5.4 Driver and electrical front-end (post-DAC analog path)	129
5.4.1 Linear gain and bandwidth limitation	129
5.4.2 Nonlinear behaviors (saturation/clipping)	130
5.4.3 Electrical noise sources and where they enter	131
5.4.4 Interconnect elements	132
5.5 Optical carrier and modulation (laser + bias + MZM)	133
5.5.1 CW laser model used in the surrogate	133
5.5.2 Bias control concept and bias-point definition	135
5.5.3 MZM transfer function currently implemented	136
5.5.4 MZM non-idealities considered for the final version	138
5.6 Channel model (TX-side propagation and noise-injection strategy)	139
5.6.1 Attenuation and measured-loss hooks	139
5.6.2 Chromatic dispersion modeling choice	140
5.6.3 PMD and other fiber impairments	141
5.6.4 ASE noise injection point and OSNR control	142

Chapter 6: SOA Preamplifier, Receiver Front-End, and Analog to Digital Converter

6.1 SOA Model in the Digital Surrogate	144
6.1.1 SOA role in IM/DD receiver chains	144
6.1.2 Gain model and carrier-dynamics abstraction	145
6.1.3 Saturation behavior and gain compression law	147
6.1.4 Pattern dependence and SOA "memory" effects	149
6.1.5 ASE noise generation in an SOA	151
6.1.6 Polarization assumptions and practical implications	153
6.1.7 Parameter choices and calibration hooks	154
6.2 SOA Operating Regimes and Design Trade-offs	155
6.2.1 Gain-noise trade-off and optimum operating point	155
6.2.2 Input power dependence and saturation-induced distortion	157
6.2.3 OSNR and eye metrics under different gain settings	159
6.2.4 EDFA vs SOA as pre-amplifiers in high-speed IM/DD	161
6.3 Photodiode and TIA Model	163
6.3.1 Photodetection, responsivity, and PD vs TIA conversion	163
6.3.2 Shot noise, dark current, and thermal/TIA noise terms	165
6.3.3 Receiver bandwidth limitation and ISI/eye closure	167
6.3.4 Complete PD+TIA front-end model tied to BER/Q/eye metrics	169
6.4 ADC Surrogate Model	171
6.4.1 Why Model the ADC in a Physics-Based Digital Surrogate	171
6.4.2 Sampling and clocking constraints	172

6.4.3 Quantization and ADC static nonlinearity	174
6.4.4 Anti-Aliasing Filters and ADC Noise-Separation Strategy	176
Chapter 7: Digital Receiver Section: DSP and Machine Learning.....	178
7.1 Receiver-DSP Overview for IM-DD PAM4	178
7.1.1 Discrete-time received-signal model	178
7.1.2 Canonical DSP chain	180
7.1.3 Where DSP interfaces the digital surrogate blocks	181
7.2 Front-End Digital Processing (Common PAM4 Practice).....	182
7.2.1 Resampling / sample-rate alignment and time-axis consistency	182
7.2.2 Timing recovery & symbol synchronization.....	183
7.2.3 DC removal/normalization/gain control	185
7.2.4 Matched filtering vs practical receiver filtering choices in IM/DD	186
7.3 The Equalizer's Role in IM-DD DSP	187
7.3.1 Why equalization is unavoidable	187
7.3.2 Equalizer as an inverse problem	188
7.3.3 Placement options	189
7.3.4 Performance metrics used to judge equalization	190
7.4 Classical Equalizers for IM-DD PAM4.....	191
7.4.1 Linear equalizers: FFE	191
7.4.2 Decision-feedback equalizer (DFE)	193
7.4.3 Adaptive algorithms.....	195
7.4.4 Nonlinear/expanded models	197
7.4.5 Sequence-based detection	198
7.5 ML-Based Equalizers for IM-DD.....	199
7.5.1 Why ML helps.....	199
7.5.2 Main families: DNN/MLP, CNN, RNN/LSTM/GRU, reservoir/ESN	201
7.5.3 Training/inference workflow for equalization.....	203
7.5.4 What "success" means.....	205
7.6 DSP Strategy for BPAM.....	206
7.6.1 BPAM: State of the art of oversampling-enabled BPAM/DD.....	206
7.6.2 Analytical basis of oversampling-enabled BPAM/DD.....	208
7.6.3 BPAM-specific processing chain and experimental requirements	210
7.6.4 Equalization implications for BPAM.....	211
7.7 Experimental Demonstration of Convolutional Neural Network Equalizer	212
7.7.1 Experimental setup	212
7.7.2 Data acquisition and preprocessing pipeline.....	214
7.7.3 Mathematical foundations of the CNN equalizer.....	216
7.7.4 CNN architecture used in this work.....	218
7.7.5 Training procedure	220
7.7.6 Inference and decision stage	221
7.7.7 Results and comparison to classical baselines	222
7.7.8 Practical discussion	224
Chapter 8: Conclusions and Perspectives	226
8.1 Executive summary of the thesis	226
8.2 Answers to the research questions	228
8.2.1 RQ1: How can a modular end-to-end surrogate preserve physically meaningful modeling.....	228
8.2.2 RQ2: How can dominant non-idealities be modeled at sample-level fidelity.....	229
8.2.3 RQ3: How can the surrogate generate ML-ready datasets.....	230
8.2.4 RQ4: How should surrogate fidelity be validated against experimental data.....	231
8.3 Achievement of the thesis objectives	232
8.3.1 O1: Architecture: define a rigorous, modular digital-surrogate framework	232
8.3.2 O2: Device-aware modeling: implement dominant non-idealities.....	233
8.3.3 O3: ML-readiness, enable structured dataset generation	234

8.3.4 O4: Validation and demonstration	235
8.4 Consolidated list of contributions	236
8.5 Limitations	237
8.6 Perspectives and future work	238
8.6.1 Near-term: full calibration and upgrade to a digital shadow	238
8.6.2 Mid-term: high-fidelity multi-path extensions	239
8.6.3 Long-term: upgrade to a full digital twin	240
Bibliography	241

List of Abbreviations

5G	Fifth Generation (mobile networks)
6G	Sixth Generation (mobile networks)
AAF	Anti-Aliasing Filter
AC	Alternating Current
ADC	Analog-to-Digital Converter
AFE	Analog Front-End
AGC	Automatic Gain Control
AI	Artificial Intelligence
APD	Avalanche Photodiode
API	Application Programming Interface
ASE	Amplified Spontaneous Emission
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BO	Bayesian Optimization
BPAM	Bipolar Pulse Amplitude Modulation
BiPAM	Bipolar Pulse Amplitude Modulation
C-band	Conventional band (1530-1565 nm)
CD	Chromatic Dispersion
CE	Circular Economy
CEAP	Circular Economy Action Plan
CH	Carrier Heating
CNN	Convolutional Neural Network
CPS	Cyber-Physical Systems
CW	Continuous Wave
CWDM	Coarse Wavelength Division Multiplexing
DAC	Digital-to-Analog Converter
DC	Direct Current
DD	Direct Detection
DFE	Decision Feedback Equalizer
DML	Directly Modulated Laser
DMT	Discrete Multi-Tone
DNL	Differential Nonlinearity
DNN	Deep Neural Network
DPD	Digital Predistortion
DSB	Double Sideband
DSP	Digital Signal Processing
DWDM	Dense Wavelength Division Multiplexing
EDFA	Erbium-Doped Fiber Amplifier
EH	Eye Height
EML	Electro-Absorption Modulated Laser
ENOB	Effective Number of Bits
ER	Extinction Ratio
ESN	Echo State Network
ETFA	Emerging Technologies and Factory Automation
EU	European Union

EVM	Error Vector Magnitude
EW	Eye Width
FEC	Forward Error Correction
FFE	Feed-Forward Equalizer
FFE/DFE	Linear equalization
FIR	Finite Impulse Response
FSE	Symbol-spaced vs fractionally-spaced FFE
FWHM	Full Width at Half Maximum
FWM	Four-wave mixing
GRU	Gated Recurrent Unit
HVAC	Heating, Ventilation, and Air Conditioning
I/O	Input/Output
ICT	Information and Communication Technology
IEEE	Institute of Electrical and Electronics Engineers
IIR	Infinite Impulse Response
IM	Intensity Modulation
IM/DD	Intensity Modulation with Direct Detection
INL	Integral Nonlinearity
INL/DNL	Static non-linearities
ISI	Inter-Symbol Interference
ITS	Intelligent Transport Systems
IoT	Internet of Things
LD	Laser Diode
LED	Light-Emitting Diode
LEZ	Low Emission Zone
LMS	Least Mean Squares
LRRM	Line-Reflect-Reflect-Match
LSB	Denote the ideal least significant bit
LSTM	Long Short-Term Memory
MAP	Maximum A Posteriori
MCMC	Markov Chain Monte Carlo
ML	Machine Learning
MLP	Multi-Layer Perceptron
MLSE	Maximum Likelihood Sequence Estimation
MMF	Multi-Mode Fiber
MMSE	Minimum Mean Square Error
MPI	IM/DD PAM links can suffer from multipath
interference	
MSO	A representative strategy is multi-symbol output
MZM	Mach-Zehnder Modulator
MaaS	Mobility as a Service
NEB	Noise-Equivalent Bandwidth
NMS	Network management system
NRZ	Non-Return-to-Zero
O-band	Original band (1260-1360 nm)
OFDM	Orthogonal Frequency Division Multiplexing
OOK	On-Off Keying
OSNR	Optical Signal-to-Noise Ratio

PAM	Pulse Amplitude Modulation
PAM4	4-level Pulse Amplitude Modulation
PD	Photodiode
PDF	Probability Density Function
PIN	P-Intrinsic-N photodiode
PMD	Polarization Mode Dispersion
PON	Passive Optical Network
PRBS	Pseudo-Random Bit Sequence
PSD	Power Spectral Density
PUMS	Piano Urbano della Mobilità Sostenibile (Urban
Sustainable Mobility Plan)	
QAM	Quadrature Amplitude Modulation
RC	The framework supports raised-cosine
REACT-EU	Recovery Assistance for Cohesion and the
Territories of Europe	
RF	Radio Frequency
RIN	Relative Intensity Noise
RL	Another approach is reinforcement learning
RLS	Recursive Least Squares
RNN	Recurrent Neural Network
RX	Receiver
RZ	Return-to-Zero
SAR	Successive Approximation Register
SDN	Software-Defined Networking
SFDR	Spurious-Free Dynamic Range
SMF	Single-Mode Fiber
SNR	Signal-to-Noise Ratio
SOA	Semiconductor Optical Amplifier
SOH	Second-Order Hold
SPM	Self-Phase Modulation
SSMF	Standard Single-Mode Fiber
THP	Tomlinson-Harashima Precoding
TIA	Transimpedance Amplifier
TX	Transmitter
UPPER	Urban Profiling Platform for Environmental
Research	
VCSEL	Vertical-Cavity Surface-Emitting Laser
VSF	Vestigial Sideband
WDM	Wavelength Division Multiplexing
WSS	Wavelength Selective Switch
XPM	Cross-Phase Modulation
ZF	Zero Forcing
ZF/MMSE	FFE/DFE viewpoints
ZOH	Zero-Order Hold

Chapter 1

Introduction

Modern society is becoming increasingly dependent on digital infrastructures, whose capacity is now constrained not only by algorithmic complexity but also by the physical limitations of communication and computation. In this environment, short-reach optical links connecting data centers, metro aggregation points, campus networks, and emerging distributed cyber-physical systems are crucial for handling increasing data traffic while maintaining energy efficiency and cost-effectiveness. Intensity-modulation and direct-detection (IM-DD) systems remain particularly relevant in this regime due to their more straightforward implementation; however, their performance is fundamentally limited by bandwidth constraints, hardware imperfections, and strong interactions between the electronic and photonic subsystems. This thesis is positioned at the intersection of *applied artificial intelligence* and *high-speed optical communications*, within a broader Industry 5.0 and circular-economy frame. The central premise is that the next step in practical optimization of IM-DD links requires moving beyond (i) overly simplified “system-level” models that hide device behavior, and beyond (ii) isolated component models that lack an end-to-end, optimization-ready integration. The thesis therefore proposes, designs, and demonstrates a *digital surrogate* of an optical communication system: an end-to-end, modular, device-aware, and machine-learning-ready representation intended to support reproducible analysis, realistic data generation, and multi-level optimization.

1.1 Background and motivation

IM-DD links have historically benefited from architectural simplicity: the optical carrier is intensity-modulated, and the receiver performs square-law detection without requiring a coherent local oscillator. This simplicity, however, entails distinct limitations. As symbol rates increase and bandwidth margins shrink, performance is dominated by finite-memory intersymbol interference (ISI) and nonlinear distortions arising from practical mixed-signal and electro-optic devices.

In practical deployments, the transmitter chain (digital processing, DAC, driver, modulator), the optical path (loss, chromatic dispersion, filtering, amplification, and amplified spontaneous emission), and the receiver front-end (photodiode/TIA, analog filtering, ADC sampling and quantization) constitute a closely integrated end-to-end channel. Therefore, understanding and optimizing performance require considering the interactions between these components, rather than treating each subsystem as an isolated block where impairments merely sum.

Simultaneously, the wider digital transformation in Industry 5.0 promotes the development of systems that are adaptable, data-driven, and resource-aware. In communications, this leads to increased interest in learning-assisted equalization, predistortion, impairment compensation, and operational optimization. However, purely black-box learning methods often lack interpretability and robustness when faced with domain shifts (i.e., differences between training and real-world conditions). Therefore, a key aim of this thesis is to enable *structured learning and optimization* by creating a simulation framework that explicitly models device non-idealities and makes them controllable, while generating detailed metadata for analysis and machine-learning procedures.

1.2 Problem statement

Despite extensive research on IM-DD transmission, a persistent methodological gap remains:

1. **Insufficient end-to-end coupling in commonly used models.** A large portion of the literature either assumes idealized mixed-signal hardware or employs “lumped” abstractions that obscure the interaction among DAC bandwidth and nonlinearities, driver behavior, Mach-Zehnder modulator transfer characteristics, optical noise loading, and receiver sampling/quantization effects.

2. **Fragmentation between high-fidelity component models and system-level studies.** Component-accurate models often do not scale to modular end-to-end studies intended for optimization across operating points; conversely, system-level studies frequently omit the very device-level knobs required for actionable optimization and hardware alignment.

3. Limitations of current learning-based approaches. Many learning-based equalizers or compensation methods improve error metrics but do not provide a principled bridge between learned behavior and physical parameters, limiting transferability and controllability in realistic scenarios.

These issues become more critical when the research goal is not only to analyze a link at a single operating point but also to optimize it across a wide parameter space, potentially including waveform design, DSP configuration, and hardware settings, while remaining consistent with real devices.

1.3 Thesis hypothesis and objectives

Hypothesis. A modular, device-aware, end-to-end digital surrogate of an IM-DD link, implemented with disciplined interfaces, strict parameterization, and structured metadata, can simultaneously:

- i* reproduce key coupled impairments more faithfully than simplified system models,
- ii* remain computationally feasible and calibratable, and
- iii* provide a practical foundation for interpretable machine-learning-assisted optimization.

Objectives. The thesis pursues the following objectives:

- **O1-Architecture:** Define a rigorous digital-surrogate architecture for IM-DD systems that enforces block modularity, avoids hard-coded assumptions, and enables reproducible experiments.
- **O2-Device-aware modeling:** Implement and integrate dominant non-idealities across transmitter, channel, and receiver, with explicit control of impairment parameters and observation points.
- **O3-ML readiness:** Design the surrogate as a data generator that produces structured metadata linking operating conditions and device settings to waveform features and performance metrics.

- **O4 - Validation and demonstration:** Demonstrate the utility of the surrogate through DSP analysis and an experimental IM-DD study, including machine-learning-based equalization for bipolar PAM.

1.4 Research questions

The thesis is guided by four research questions:

- **RQ1:** How can an end-to-end, mixed-signal digital surrogate for IM-DD PAM/bipolar-PAM links be designed so that physically meaningful coupling across transmitter–channel–receiver is preserved within a modular framework?
- **RQ2:** How can dominant device non-idealities be modeled at sample-level fidelity while keeping the surrogate computationally feasible, parameterized, and calibratable against experiments?
- **RQ3:** How can the surrogate generate structured, ML-ready datasets and metadata that enable controllable and interpretable optimization (DSP/ML, hybrid, and end-to-end)?
- **RQ4:** How should surrogate fidelity and optimization outcomes be validated against measurements, and how can mismatch/uncertainty be quantified to mitigate domain shift across operating conditions and hardware variability?

1.5 Methodological approach

To address these questions, the thesis adopts an engineering methodology centered on disciplined modularity and hardware awareness, organized along three axes.

1.5.1 End-to-end modular architecture

The surrogate is structured as a chain of processing blocks with explicit domain transitions (digital $\rightarrow$ electrical $\rightarrow$ optical $\rightarrow$ electrical $\rightarrow$ digital). A unified interface contract is used so each block can be enabled, disabled, replaced, or calibrated independently while maintaining end-to-end coherence. The framework is parameter-driven: simulation assumptions, device settings, and execution modes

are controlled through a centralized parameter hierarchy, supporting reproducibility and systematic sweeps.

1.5.2 Device-aware impairment modeling

The surrogate models focus on impairments that are both practically dominant and strongly coupled end-to-end. On the transmitter side, this includes realistic DAC behavior (sampling-rate alignment, pre-emphasis, quantization conventions, static nonlinearities such as INL/DNL, reconstruction effects, and bandwidth), driver limitations, and electro-optic modulation nonlinearity. In the optical path, fiber-induced dispersion and noise loading mechanisms are represented in forms consistent with IM-DD evaluation practices. On the receiver side, the front end and ADC chain are modeled with explicit attention to bandwidth, noise, sampling, and quantization effects that shape the discrete-time channel observed by the DSP.

1.5.3 ML readiness and validation strategy

The surrogate is designed not only to produce a waveform, but to produce *context*: structured metadata describing operating points, impairment parameters, and observation locations. This enables controlled dataset generation for learning-based equalization or optimization. Validation is treated as an explicit design requirement: the framework supports internal consistency checks and comparison against experimental measurements, and it is structured to make sources of mismatch explicit rather than implicit.

1.6 Main contributions

The main contributions of this thesis can be summarized as follows:

1. **A digital-surrogate framework for IM-DD links** that is end-to-end, modular, parameterized, and aligned with the needs of data-driven optimization.
2. **A structured architecture and runtime discipline** for mixed-signal simulation, including standardized block interfaces, explicit domain labeling, and mechanisms for consistent time/sampling management and visualization-ready metadata.

3. **Device-aware transmitter/channel/receiver prototypes** that expose hardware non-idealities as controllable parameters, enabling systematic exploration of coupled impairment regimes rather than isolated effects.

4. **A receiver DSP perspective** that treats the IM-DD link as a finite-memory (often nonlinear) discrete-time channel, motivating equalization and robust decision strategies at high symbol rates.

5. **An experimental demonstration and ML equalization case study** on bipolar PAM, illustrating how oversampling-based observation and learning-based equalization can recover information that is not accessible under naive square-law detection assumptions.

1.7 Thesis organization

The remainder of the thesis is organized as follows.

- **Chapter 2** frames the work within the broader context of Industry 5.0 and circular-economy motivations, and connects the technical development of optical communication surrogates to an industrial mission and application perspective.
- **Chapter 3** reviews the state of the art in short-reach IM-DD systems, digital-twin/surrogate concepts, and machine-learning practice, and identifies the methodological gap addressed by this thesis.
- **Chapter 4** presents the proposed digital-surrogate architecture, including design principles, modular interfaces, and the rationale for parameterization and metadata-driven workflows.
- **Chapters 5 and 6** describe the transmitter, channel, and receiver front-end modeling, emphasizing how device non-idealities are represented as explicit knobs and how intermediate observation points support analysis and optimization.
- **Chapter 7** focuses on the digital receiver: discrete-time modeling, DSP requirements, and a detailed experimental study including a machine-learning-based equalization case study for bipolar PAM.

- **Chapter 8** concludes the thesis by summarizing the achieved objectives, discussing limitations and calibration considerations, and outlining perspectives for extending the surrogate toward broader formats, tighter experimental alignment, and closed-loop optimization.

1.8 Notation and scope of applicability

Unless otherwise stated, the thesis concentrates on short-reach IM-DD links and PAM-family(including Bipolar Optical PAM) modulation formats, including bipolar variants enabled through suitable sampling and DSP. The term digital surrogate refers to an end-to-end, parameterized representation of the physical link designed for analysis, data generation, and optimization. Although the prototypes and demonstrations focus on IM-DD, the architectural principles are designed for extensibility, enabling future support for additional modulation formats, device models, and experimental integration.

Chapter 2

Industry 5.0 and Smart Urban Mobility: Circular Economy Perspectives and the Role of Optical Communication

2.1 Purpose and Position in the Thesis

2.1.1 Motivation and scope of the chapter

The previous chapter introduced the central technical theme of this research: the development of a high-fidelity, modular digital surrogate for an intensity-modulated direct-detection (IM/DD) optical communication link, designed to capture realistic device non-idealities and to support both classical and machine-learning-based signal processing. That introduction was deliberately focused on the link level and on the methodological challenges of modeling high-speed optical transceivers.

This chapter broadens the perspective. Its purpose is to situate the thesis within the broader industrial and societal context in which such an optical system will ultimately operate. In particular, the chapter connects four layers:

- The emerging Industry 5.0 paradigm,
- The principles of the circular economy,
- The development of smart cities, with a focus on smart mobility and traffic management, and
- The enabling role of optical communication infrastructures and their accurate digital surrogates.

Modern metropolitan areas increasingly rely on dense sensing, large-scale data analytics, and automated decision support to manage traffic flows, integrate different modes of transport, and support safety-critical services. These capabilities are typically grouped under the umbrella of *smart mobility* and *intelligent transport systems*, and they are core components of broader smart-city strategies. At the same

time, European policy frameworks¹ encourage industry and public administrations to move towards human-centric, sustainable, and resilient models of development, often articulated through Industry 5.0² and circular-economy³ agendas.

Within this landscape, high-capacity optical links, especially short-reach IM/DD connections inside and between data centers and edge-computing nodes, provide the digital nervous system over which mobility data are collected, processed, and acted upon. The quality, efficiency, and robustness of these optical infrastructures therefore have indirect but concrete implications for the performance and sustainability of smart-city services.

The scope of this chapter is to make these connections explicit. Rather than presenting new physical or algorithmic results, it offers a conceptual mapping between:

- The macro-level objectives of Industry 5.0 and the circular economy,
- The design and governance of smart cities and smart mobility systems, and
- The micro-level technical work on optical communication and digital surrogate modeling is developed in the rest of the thesis.

This kind of mapping provides the strategic justification for investing in realistic, hardware-aware optical system models: such models are not only scientifically interesting but also relevant to the design of digital infrastructures that support sustainable, human-centric urban services.

2.1.2 Relation to the thesis and doctoral programme

The research presented in this thesis is conducted within a doctoral programme explicitly dedicated to big data and artificial intelligence for Industry 5.0 and the circular economy, and is supported by an industrial partner active in the development of digital and smart-city solutions. While the technical core of the thesis is firmly rooted in optical communication and signal processing, the

¹ K. Breque, L. De Nul, A. Petridis, *Industry 5.0 - Towards a sustainable, human-centric and resilient European industry*, European Commission, Brussels 2021.

² *Ibidem*.

³ European Commission, *A new Circular Economy Action Plan - For a cleaner and more competitive Europe*, European Commission, Brussels 2020.

programme framework and the industrial sponsorship both point towards a broader application domain: smart, sustainable cities and mobility services.

This chapter translates that institutional framing into a scientific and engineering narrative. It clarifies why a seemingly specialized topic such as “end-to-end modeling of IM/DD optical links and their non-ideal components” is directly relevant to:

- the data and communication backbones of smart-city platforms,
- the operational needs of traffic management and smart mobility systems, and
- policy agendas in which resource efficiency, system resilience, and human-centric design are central.

In doing so, the chapter serves two complementary functions:

- It anchors the thesis contributions in a realistic use-case family of urban mobility and traffic management in large metropolitan areas, such as the Rome context, which will be used as a running example.
- It provides a conceptual bridge between the high-level language of policy and industrial roadmaps (Industry 5.0, circular economy, smart city) and the more detailed technical chapters that follow (system architecture, device models, DSP and ML-based equalization, experimental validation).

The following chapters return to a more technical focus; the present chapter should be read as a contextual chapter that explains why the optical digital surrogate matters to stakeholders beyond the laboratory.

2.1.3 Organization of the chapter

The organization of the remainder of this chapter is as follows.

- Section 2.2 introduces the conceptual framework of Industry 5.0 and the circular economy. It reviews their main principles, highlights their differences with earlier Industry 4.0 narratives, and discusses how they jointly emphasize human-centricity, sustainability, and resilience in industrial and infrastructural systems.
- Section 2.3 examines smart cities, with particular emphasis on smart mobility and urban traffic management. It presents standard definitions and

dimensions of smart cities, outlines the functional architecture of modern traffic management and mobility platforms, and uses the metropolitan area of Rome as an illustrative example of the challenges and opportunities involved.

- Section 2.4 focuses on the role of digital and communication infrastructures as the backbone of smart cities. It describes how high-capacity wired and wireless networks, data centres, edge computing nodes, and digital platforms support real-time sensing, data fusion, and control in smart mobility scenarios. It highlights the role of short-reach IM/DD optical links within this architecture.
- Section 2.5 introduces the concept of a digital surrogate for an optical communication system and situates the surrogate developed in this thesis within the broader family of digital-twin approaches used in smart infrastructure. It explains how such a surrogate can contribute to the goals of Industry 5.0 and the circular economy by enabling more efficient design, operation, and evolution of the underlying communication infrastructure.

In this way, the chapter provides a top-down view, ranging from societal and industrial objectives to the optical communication layer, thereby preparing the ground for the detailed bottom-up modeling and analysis that constitute the core scientific contribution of the thesis.

2.2 Industry 5.0 and Circular Economy

2.2.1 From Industry 4.0 to Industry 5.0

Over the last decade, Industry 4.0 has become a dominant paradigm for describing the digital transformation of manufacturing and industrial systems. It is commonly associated with the integration of cyber-physical systems (CPS), the Internet of Things (IoT), cloud computing, big data analytics, and advanced automation into production processes, thereby enabling real-time monitoring, data-

driven decision-making, and “smart factories”^{4,5}. Within this framework, the focus is primarily on efficiency, flexibility, and productivity gains achieved through the widespread use of digital technologies and interconnected machinery across the value chain⁶.

The more recent notion of Industry 5.0, promoted in particular by the European Commission, is not intended as a simple chronological “next step” after Industry 4.0, but as a complementary re-orientation of industrial development⁷.

Whereas Industry 4.0 emphasises technological capabilities and automation, Industry 5.0 explicitly foregrounds three normative pillars:

- Human-centricity: placing the well-being, skills, and empowerment of workers and citizens at the centre of industrial transformation;
- Sustainability: aligning industrial activity with planetary boundaries and long-term environmental goals;
- Resilience: strengthening the ability of industrial and infrastructural systems to withstand and adapt to shocks, disruptions, and crises⁸.

In this perspective, digital technologies such as artificial intelligence, advanced robotics, cyber-physical systems, and high-capacity communication networks are no longer ends in themselves but means to support human, social, and environmental objectives. The Industry 5.0 policy brief explicitly frames industry as a “resilient provider of prosperity” that creates value for a broad set of stakeholders, rather than focusing narrowly on shareholder value and short-term productivity gains⁹. Recent academic work echoes this view, characterizing

⁴ A. Khan, K. Turowski, “A Preliminary Study on Industry 4.0,” *Journal of Industrial and Intelligent Information*, 4(3), 2016.

⁵ Javaid *et al.*, “An integrated outlook of cyber-physical systems for Industry 4.0, Topical practices, architecture, and applications” *Green Technologies and Sustainability*, 1(1), 100001, 2023.

⁶ F. Yang *et al.*, “Industry 4.0: A revolution that requires technology and national strategies,” *Complex & Intelligent Systems*, 2021.

⁷ K. Breque, L. De Nul, A. Petridis, *Industry 5.0 - Towards a sustainable, human-centric and resilient European industry*, European Commission, Brussels 2021.

⁸ European Commission, *Industry 5.0 - Research and Innovation for a Sustainable, Human-Centric and Resilient European Industry*, Research & Innovation Policy brief, 2021.

⁹ K. Breque, L. De Nul, A. Petridis, *Industry 5.0 - Towards...*, Brussels 2021, see note 6.

Industry 5.0 as a human-centric paradigm for sustainable and resilient industrial transformation, in which technological innovation is systematically oriented toward the quality of work, environmental performance, and societal resilience¹⁰.

For this thesis, two aspects of this transition from Industry 4.0 to Industry 5.0 are particularly relevant. First, the human-centric and sustainable orientation motivates greater attention to the energy efficiency, robustness, and lifecycle impacts of digital and communication infrastructures themselves, including data centres and optical transport networks. Second, the emphasis on resilience and stakeholder value aligns naturally with critical infrastructures such as smart-city platforms and intelligent transport systems, where failures and inefficiencies can have immediate societal consequences. In this sense, Industry 5.0 provides the macro-level framework within which the detailed modeling and optimisation of optical communication systems, presented in the technical chapters of this thesis, acquire strategic relevance.

2.2.2 Circular economy principles and policy background

The concept of the circular economy (CE) has gained significant traction over the last decade as an alternative to the traditional linear “take-make-dispose” industrial model. In contrast to this linear paradigm, the circular economy is typically defined as a regenerative system, in which resource input and waste, emissions, and energy leakage are minimised by slowing, closing, and narrowing material and energy loops through strategies such as long-lasting design, maintenance, repair, reuse, remanufacturing, refurbishing, and recycling¹¹. At the same time, numerous partially overlapping definitions exist in the literature: a well-known review of 114 definitions shows that CE is most frequently described as a combination of “reduce-reuse-recycle” activities, while its systemic and

¹⁰ J. Brodny M. Tuta “Assessing Human-Centric, Resilient, Sustainable and Smart Transformation in the EU-27 Countries: A New Balanced Composite Index Method”, *Journal of Open Innovation: Technology, Market, and Complexity*, 100685, 2025.

¹¹ M. Geissdoerfer, P. Savaget, N. M. P. Bocken, E. J. Hultink, “The Circular Economy - A new sustainability paradigm?,” *Journal of Cleaner Production*, 2017.

transformational nature is sometimes underemphasised¹². Despite this definitional plurality, most accounts agree that CE aims to decouple economic activity from finite resource consumption and to integrate economic, environmental, and, increasingly, social objectives.

Influential organisations such as the Ellen MacArthur Foundation have played a central role in popularising and operationalising the CE concept. Their early reports describe the circular economy as an industrial system that is restorative or regenerative by intention and design, and propose practical frameworks such as the ReSOLVE set of actions (Regenerate, Share, Optimise, Loop, Virtualise, Exchange), to guide businesses and policymakers in developing circular strategies and business models¹³. These frameworks emphasise that circularity is not limited to end-of-pipe recycling but encompasses interventions across all stages of the product and infrastructure life cycle, from design and material selection to new service-based models, reverse logistics, and secondary-material markets. More recent synthesis work positions the circular economy as a specific approach to sustainability, in which economic performance, social inclusiveness, and environmental resilience are balanced through the design of regenerative and restorative systems¹⁴.

In Europe, the circular economy has been institutionalised as a strategic policy priority. The European Union adopted its first Circular Economy Action Plan in 2015, focusing on improved product design, higher recycling rates, and more efficient waste management, and then reinforced and broadened this agenda with the new Circular Economy Action Plan (CEAP) of 2020, identified as one of the main building blocks of the European Green Deal¹⁵. The 2020 CEAP promotes a comprehensive life-cycle approach that addresses how products are designed,

¹² J. Kirchherr, D. Reike, M. Hekkert, “Conceptualizing the circular economy: An analysis of 114 definitions,” *Resources, Conservation and Recycling*, 2017.

¹³ Ellen MacArthur Foundation, *Towards the Circular Economy Vol. 1: An economic and business rationale for an accelerated transition*, 2013.

¹⁴ M. Geissdoerfer *et al.*, “The Circular Economy - A new sustainability paradigm?,” 2017.

¹⁵ European Commission, *Closing the loop - An EU action plan for the Circular Economy*, COM(2015) 614 final, Brussels 2015;.

produced, used, and recovered, with particular attention to resource-intensive sectors such as electronics, batteries, construction, packaging, plastics, and mobility. The overarching objective is to keep resources in the EU economy for as long as possible, reduce waste, and enable a cleaner, more competitive Europe.

Cities and regions are recognised as key arenas for implementing circular-economy strategies, because they concentrate population, economic activity, infrastructure, and material flows. Policy-oriented analyses from organisations such as the OECD highlight that applying CE principles at the urban scale involves rethinking buildings, energy and water systems, mobility, waste management, and public spaces to avoid “linear lock-in” in infrastructure and to foster closed-loop resource cycles¹⁶. Environmental research on the circular economy in cities shows that municipal initiatives often focus on sectors such as construction, waste, water, and food systems, but also stresses the need to better quantify environmental impacts and align local practices with systemic circularity objectives¹⁷. In this thesis, the urban and infrastructural dimension is significant because it links circular-economy policy goals to the physical networks and assets, including digital and communication infrastructures, that underpin smart-city and smart-mobility services. The following subsection examines how these CE objectives intersect with the Industry 5.0 paradigm and how they together inform the transformation of urban systems.

2.2.3 How Industry 5.0 and the circular economy reinforce each other

Industry 5.0 and the circular economy can be viewed as two mutually reinforcing agendas for re-thinking production, infrastructures, and service systems beyond the linear “take-make-dispose” logic¹⁸. While Industry 5.0 provides a

¹⁶ European Commission, *A new Circular Economy Action Plan - For a cleaner and more competitive Europe*, COM(2020) 98 final, Brussels 2020.

¹⁷ A. Petit-Boix, S. Leipold, “Circular economy in cities: Reviewing how environmental research aligns with local practices,” *Journal of Cleaner Production*, 2018.

¹⁸ A. Rejeb, K. Rejeb, J. G. Keogh, E. Süle, “When Industry 5.0 Meets the Circular Economy: A Systematic Literature Review,” *Circular Economy and Sustainability*, vol. 5, no. 4, pp. 2621-2652, 2025.

paradigm for how industrial systems should be designed, human-centric, sustainable, and resilient, the circular economy offers a paradigm for how resources should flow, via slowed, closed, and narrowed loops that maximise value retention and minimise waste¹⁹. In many policy and academic contributions, circular-economy strategies are explicitly framed as a way to operationalise the sustainability pillar of Industry 5.0 at the level of value chains, territories, and ecosystems rather than within individual firms only¹⁸.

From this perspective, Industry 5.0 can be interpreted as a direction-setting framework that orients innovation and investment (e.g., towards human-centric AI, green digital infrastructures, skills, and resilience). In contrast, the circular economy offers a toolbox of concrete strategies and business models (product-service systems, remanufacturing, industrial symbiosis, sharing models, regenerative design) that translate those directions into practice²⁰. Systematic reviews and empirical studies increasingly show that firms and regions that develop strong Industry 5.0 technological capabilities, namely AI, big data and analytics, IoT, robotics, and blockchain, are better positioned to implement circular practices and to improve environmental, social, and economic performance²¹.

In turn, circular-economy thinking pushes Industry 5.0 beyond an incremental “efficiency” agenda. The R-strategies (refuse, reduce, reuse, repair, refurbish, remanufacture, repurpose, recycle, recover) provide a structured vocabulary for redesigning products, processes, and infrastructure to achieve long lifetimes, modularity, and material recovery¹⁹. Recent decision-making frameworks, such as “Waste Management 5.0”, explicitly combine circular-economy indicators (e.g., 5R performance) with Industry-5.0-type technological enablers (smart sensors, AI-driven optimisation, cyber-physical systems) to prioritise interventions that both

¹⁹ S. Dixon-Declève, P. A. Balland *et al.*, *Industry 5.0: A Transformative Vision for Europe - Governing Systemic Transformations Towards a Sustainable Industry*, ESIR Policy Brief No. 3, European Commission, Directorate-General for Research and Innovation, Brussels, 2022.

²⁰ S. Nayeri, Z. Sazvar, E. B. Tirkolaei, “Towards Waste Management 5.0: A Nexus Between Circular Economy and Industry 5.0 Dimensions,” *Waste Management*, vol. 203, 114830, 2025.

²¹ M. N. Shafique, U. Adeel, A. Rashid, “The Synergy Between Industry 5.0 and Circular Economy for Sustainable Performance in the Chinese Manufacturing Industry,” *Sustainability*, vol. 16, no. 22, 9952, 2024.

reduce environmental impact and exploit advanced technologies in a socially acceptable manner¹⁹.

Several studies highlight that this mutual reinforcement is not automatic but depends on how Industry 5.0 technologies are deployed²⁰. If AI, robotics, and hyper-automation are used mainly to accelerate throughput and labour substitution, they can lock systems into high-consumption, resource-intensive trajectories that undermine circular-economy goals. Conversely, when guided by circular-economy principles and broader sustainability objectives, Industry 5.0 capabilities can enable new forms of value creation, such as data-driven predictive maintenance, sharing platforms, product-service systems, and local repair ecosystems, that simultaneously improve resilience, reduce resource use, and support decent work²².

Conceptual frameworks such as the “Industry 5.0 framework” explicitly argue for aligning business models and supply chains with a new triple bottom line, *resilient value creation, human well-being, and sustainable society*, within which circular-economy strategies are core mechanisms rather than optional add-ons²³.

For this thesis, it is precisely this synergistic view that matters. The digital surrogate developed in the following chapters focuses on high-capacity optical communication links and their realistic non-idealities. However, the surrounding narrative is an Industry 5.0 narrative: a human-centric use of AI and digital twins to enhance infrastructure resilience and sustainability. When short-reach optical networks are deployed to enable smart, shared, and efficiently managed urban mobility services, they contribute to circular-economy objectives by reducing congestion, energy use, and material over-dimensioning in transport systems. In this sense, the digital surrogate is not only a tool for physical-layer optimisation but also a building block within a broader socio-technical transition in which Industry 5.0 and the circular economy mutually reinforce one another to support sustainable, resilient, and human-centric smart cities.

²² S. Khan *et al.*, “The Role of Industry 5.0 in Achieving Circular Economy Goals: Integrating Smart Technologies for Sustainable Supply Chains,” *TEM Journal*, vol. 14, no. 4, pp. 3144-3154, 2025.

²³ D. Ivanov, “The Industry 5.0 Framework: Viability-Based Integration of the Resilience, Sustainability, and Human-Centricity Perspectives,” *International Journal of Production Research*, vol. 61, no. 5, pp. 1683-1695, 2023.

2.3 Smart Cities and Smart Urban Mobility

2.3.1 Definitions and dimensions of a smart city

The term *smart city* has become a central but contested concept in urban studies, information systems, and policy documents. Recent reviews show that there is no single, universally accepted definition; instead, smart cities are typically described as urban systems that leverage digital technologies, data, and integrated ICT infrastructures to improve the efficiency, sustainability, and quality of urban services, while enhancing quality of life and governance²⁴. At the same time, many authors emphasise that smart cities are not merely *technology-intensive* cities, but multidimensional socio-technical systems in which technology, institutions, and communities co-evolve²⁵.

From an information systems perspective, smart cities are often defined as cities that use advanced ICT and data analytics to support integrated, context-aware decision-making across domains such as mobility, the environment, the economy, public services, and governance²⁶. This view highlights the role of platforms, data flows, and interoperability among heterogeneous systems (e.g., traffic management, energy grids, public transport, emergency services) as key enablers of “smartness”. In parallel, structured literature reviews underline that smart-city definitions usually combine three types of elements: (i) enabling technologies (IoT, sensors, communication networks, AI, data analytics), (ii) application domains

²⁴ J. S. Gracias, G. S. Parnell, E. Specking, E. A. Pohl, and R. Buchanan, “Smart Cities-A Structured Literature Review,” *Smart Cities*, vol. 6, no. 4, pp. 1719-1743, 2023.

²⁵ T. Yigitcanlar, M. Kamruzzaman, L. Buys *et al.*, “Understanding ‘Smart Cities’: Intertwining Development Drivers with Desired Outcomes in a Multidimensional Framework,” *Cities*, vol. 81, pp. 145-160, 2018.

²⁶ E. Ismagilova, L. Hughes, Y. K. Dwivedi, and K. R. Raman, “Smart Cities: Advances in Research-An Information Systems Perspective,” *International Journal of Information Management*, vol. 47, pp. 88-100, 2019.

(mobility, environment, governance, economy, people, living), and (iii) desired outcomes, such as sustainability, resilience, liveability, and inclusive growth²⁷.

A widely adopted operational framework is the six-dimensional model proposed by Giffinger and colleagues, which characterises a smart city through the dimensions of *Smart Economy*, *Smart People*, *Smart Governance*, *Smart Mobility*, *Smart Environment*, and *Smart Living*²⁸. This model, initially developed for ranking European medium-sized cities, has become a de facto reference in both academic and policy work, because it makes explicit that “smartness” combines economic competitiveness, human and social capital, environmental performance, mobility and transport, governance quality, and quality of life²⁸. Later conceptual work, such as the multidimensional framework by Yigitcanlar *et al.*, intertwines *development drivers* (technology, community, policy) with *desired outcomes* (productivity, sustainability, accessibility, wellbeing, liveability, governance), reinforcing the idea that smart cities must be assessed along multiple, interdependent dimensions rather than by technology adoption alone²⁵.

In more recent reviews, smart cities are increasingly framed as data-intensive and governance-intensive configurations²⁶. Digital infrastructures (sensors, communication networks, data centres, edge/cloud platforms) and algorithmic systems (analytics, AI, decision-support tools) are seen as necessary but not sufficient conditions; equally important are institutional capabilities, regulatory frameworks, and citizen engagement mechanisms that ensure that data and technology are used to pursue public values such as sustainability, equity, and resilience²⁶. This is particularly relevant for the *Smart Mobility* dimension, which directly concerns the design and management of urban transport systems, and for which technological integration must be aligned with broader policy objectives (e.g., emissions reduction, accessibility, safety).

In the context of this thesis, smart cities are therefore understood as multidimensional, socio-technical systems that use digital technologies, data, and

²⁷ *Ibidem*.

²⁸ R. Giffinger, C. Fertner, H. Kramar, R. Kalasek, N. Pichler-Milanović, and E. J. Meijers, *Smart Cities. Ranking of European Medium-Sized Cities. Final Report*, Vienna University of Technology, 2007.

AI to coordinate complex infrastructures and services, including transport, towards sustainability, resilience, and improved quality of life. Within this broader picture, the *Smart Mobility* dimension and its underlying digital infrastructures (sensors, communication networks, data platforms) form the main point of contact between the conceptual smart-city literature and the physical-layer work developed in the following chapters. The digital surrogate of an optical communication system presented later can be seen as a building block within the broader ecosystem of smart-city infrastructures that support high-capacity, low-latency communication for urban mobility and traffic management services.

2.3.2 Urban traffic management as a smart-city “nervous system.”

In the smart-city literature, urban mobility is repeatedly described as a cornerstone or even a backbone of the overall smart-city strategy: if people and goods cannot move efficiently, sustainably, and safely, many of the other “smart” functions, such as digital services, economic activity, and emergency response, will lose their effectiveness²⁹. Urban traffic management is therefore not just a technical subsystem for coordinating traffic lights or enforcing speed limits; it is one of the primary control layers through which the city shapes how flows of people, vehicles, and information traverse the urban space. In this sense, several authors and industrial reports liken the combination of sensing, communication networks, and control centers in a smart city to a “nervous system”, with mobility-related infrastructures as some of its most critical pathways³⁰.

From a systems perspective, contemporary urban traffic management systems (UTMS) are typically decomposed into three tightly coupled functional layers: a data-collection layer (sensors, detectors, connected vehicles), a traffic management layer (control algorithms, decision-support tools, operations centers), and a safety and enforcement layer (incident detection, compliance monitoring, emergency

²⁹ C. Biyik *et al.*, “Smart mobility adoption: A review of the literature,” *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 7, no. 2, p. 146, 2021

³⁰ Ismagilova, L. Hughes, Y. K. Dwivedi, and K. Raman, “Smart cities: Advances in research-An information systems perspective,” *International Journal of Information Management*, 47, 88–100, 2019.

response)³¹. The data-collection layer aggregates heterogeneous streams from inductive loops, CCTV, radar, LIDAR, floating-car data, cooperative ITS messages, and environmental sensors. The management layer uses these data to estimate current traffic states, predict near-future conditions, and compute control actions (e.g., signal timings, variable speed limits). In contrast, the safety layer translates both real-time and historical insights into interventions that reduce accidents, protect vulnerable road users, and prioritize emergency vehicles.

In the context of smart mobility, this architecture evolves into a networked, cyber-physical system that integrates road traffic, public transport, logistics, and active modes (walking, cycling) into a single control problem³². Intelligent Transport Systems (ITS) support this integration through adaptive traffic signal control, coordinated corridor management, dynamic route guidance, transit signal priority, smart parking, and demand-responsive transport³³. Rather than being optimized in isolation, these functions are increasingly designed as components of a multimodal management layer that balances private cars, shared mobility services, public transport, and micromobility in accordance with policy goals such as congestion reduction, emission targets, and modal shift objectives.

This “nervous system” metaphor becomes particularly clear when looking at traffic management centers (TMCs) and urban data platforms. These facilities act as the “brain” of the mobility subsystem: ingesting real-time data from the “senses” (distributed sensors and connected devices), processing them through analytics and AI models, and sending control commands back to the field, such as updated signal plans, dynamic lane assignments, or emergency pre-emption for ambulances³⁴. High-capacity communication networks, including optical backbones and edge

³¹ A. Ait Ouallane et al., “Fusion of engineering insights and emerging trends in traffic management systems,” *Simulation Modelling Practice and Theory*, 120, 102624, 2022.

³² P. Ribeiro, G. Dias, and P. Pereira, “Transport systems and mobility for smart cities,” *Applied System Innovation*, 4(3), 61, 2021.

³³ R. Mangiaracina, A. Perego, G. Salvadori, and A. Tumino, “A comprehensive view of intelligent transport systems for urban smart mobility,” *International Journal of Logistics Research and Applications*, 20(1), 39-52, 2017.

³⁴ F. Alanazi and M. Alenezi, “A framework for integrating intelligent transportation systems with smart city infrastructure,” *Journal of Infrastructure, Policy and Development*, 8(5), 2024.

connectivity, form the signal pathways that allow this loop to operate within tight temporal constraints. If these pathways are congested, unreliable, or too slow, the nervous system lapses from proactive to merely reactive behavior, undermining the smart-city promise of anticipatory, data-driven governance.

Recent reviews on smart mobility and ITS highlight that, when these layers are well integrated, innovative urban traffic management can yield measurable benefits: reduced congestion and travel times, lower fuel consumption and emissions, improved safety, and more reliable public transport³⁵. At the same time, they underline persistent challenges: institutional fragmentation between transport agencies and IT departments, legacy systems that limit interoperability, cybersecurity and privacy risks, and difficulties in aligning real-time optimization with long-term planning and equity goals³⁶. These challenges are especially pronounced in large metropolitan areas, where multiple jurisdictions, operators, and infrastructure owners must coordinate over a shared network.

Within this thesis, urban traffic management is therefore viewed as a domain-specific instantiation of the broader smart-city nervous system: the place where high-frequency, safety-critical control loops intersect with long-term sustainability and quality-of-life objectives. The communication infrastructure that connects sensors, control centers, and roadside equipment is not treated as a passive background, but as a performance-limiting component whose latency, capacity, and reliability directly affect how finely the city can sense, predict, and actuate on mobility flows. The digital surrogate of an optical communication system developed in the following chapters is conceived precisely as a tool to understand and optimize this “neural layer” in realistic conditions, so that future smart-city traffic-management schemes, particularly in dense metropolitan areas, can rely on communication links whose behavior is predictable, tunable, and robust under real-world workloads.

³⁵ P. Munhoz et al., “*Smart mobility: The main drivers for increasing the intelligence of urban mobility*,” *Sustainability*, 12(24), 10675, 2020.

³⁶ D. Mitieka, R. Luke, H. Twinomurinzi, and J. Mageto, “*Smart mobility in urban areas: A bibliometric review and research agenda*,” *Sustainability*, 15(8), 6754, 2023.

2.3.3 Data, AI, and digital infrastructure requirements in smart mobility

Smart urban mobility is structurally data-intensive. Modern transport systems rely on continuous streams of information from fixed and mobile sensors, including inductive loops, cameras, radar, LiDAR, environmental stations, GPS traces, automatic passenger counting, and connected vehicles³⁷. At the city scale, these heterogeneous sources generate high-volume, high-velocity data that must be ingested, cleaned, and fused in near-real-time to support traffic-state estimation, incident detection, and multimodal coordination. Recent systematic reviews on IoT and sensing in smart cities emphasize that mobility is one of the domains with the densest sensor deployments and the strongest requirements for temporal resolution and reliability, since many control decisions (e.g., adaptive signal timing) are constrained by sub-minute or even sub-second reaction times³⁸.

On top of this sensing layer, AI and advanced analytics have become central to smart mobility. In the context of intelligent transport systems (ITS), AI methods are used for traffic flow prediction, congestion hotspot identification, incident detection, travel time estimation, demand forecasting, and dynamic routing³⁹. Literature surveys on AI in ITS and adaptive traffic control report a rapid proliferation of deep learning architectures (CNNs, RNNs, graph neural networks) for short-term traffic prediction, reinforcement learning for adaptive signal control, and hybrid models that combine physics-based transport models with data-driven components⁴⁰. The goal is to transition from static, schedule-based operations to anticipatory control, in which the system can forecast near-future conditions and proactively adjust signal plans, speed limits, or service frequencies. In practice, this requires not only algorithmic accuracy but also a communication and computing

³⁷ D. Oladimeji *et al.*, “Smart transportation: An overview of technologies and applications,” *Sensors*, vol. 23, no. 8, p. 3880, 2023.

³⁸ F. Zeng, C. Pang, and H. Tang, “Sensors on Internet of Things systems for the sustainable development of smart cities: A systematic literature review,” *Sensors*, vol. 24, no. 7, p. 2074, 2024

³⁹ C. Gheorghe and A. Soica, “Revolutionizing urban mobility: A systematic review of AI, IoT, and predictive analytics in adaptive traffic control systems for road networks,” *Electronics*, vol. 14, no. 4, p. 719, 2025

⁴⁰ A. Kumar *et al.*, “Navigating urban mobility: A review of AI-driven traffic flow management in smart cities,” in *Proc. ICRITO 2024*.

substrate capable of executing inference and control loops within tight latency bounds.

These AI-enhanced applications are enabled by a multilayer digital infrastructure spanning field devices, edge nodes, and centralized platforms. At the communication layer, smart mobility depends on a mix of short-range and wide-area technologies: vehicle-to-everything (V2X) links, Wi-Fi and cellular networks at the access edge, and high-capacity wired backbones, often optical fiber, for aggregation and backhaul⁴¹. Surveys of innovative transportation architectures typically highlight the importance of reliable, low-latency connectivity between sensors, roadside units, and control centers, with 5G and optical networks playing complementary roles: wireless links provide flexible coverage and support for moving users, while optical cores offer the deterministic capacity required to transport large volumes of sensor, video, and control data without becoming a bottleneck⁴². From a systems viewpoint, the communication fabric is not merely a conduit but a co-designed component of the overall control system: its latency, jitter, and resilience directly shape which AI-driven control policies can be deployed in practice.

Above the network layer, cloud and edge computing platforms host the data-management and analytics functions. Edge nodes (e.g., at intersections or depots) can perform local preprocessing, privacy filtering, or lightweight inference. At the same time, regional and central data centers run more intensive analytics, model training, and historical analysis⁴³. Data platforms integrate real-time streams with static datasets (network topology, timetables, land use, socio-economic indicators) and expose them to decision-support tools, dashboards, and optimization engines. Reviews of data-driven smart-city applications stress that governance issues, including data ownership, interoperability, standard APIs, security, and privacy, are

⁴¹ F. Alanazi and M. Alenezi, “A framework for integrating intelligent transportation systems with smart city infrastructure,” *Journal of Infrastructure, Policy and Development*, vol. 8, no. 5, 2024.

⁴² *Ibidem*

⁴³ Y. Kaluarachchi, “Implementing data-driven smart city applications for future cities,” *Smart Cities*, vol. 5, no. 2, pp. 350-367, 2022

as critical as the raw computing power, particularly when multiple agencies and private operators contribute data to a shared mobility ecosystem⁴⁴.

In recent years, these data, AI, and digital infrastructure components have been increasingly integrated into *mobility digital-twin ecosystems*. Mobility digital twins aim to maintain a live, bidirectional link between the physical transport system and a high-fidelity virtual model, updated with real-time data and capable of simulating interventions before they are applied in the field⁴⁵. Case studies and conceptual frameworks describe how such twins can support sustainable urban mobility planning, operational traffic management, and resilience analysis by linking sensing, simulation, optimization, and stakeholder interaction into a coherent loop⁴⁶. For our purposes, an important point is that digital twins expose not only macroscopic transport indicators (flows, speeds, emissions) but also implicit requirements on the underlying digital infrastructure: maintaining a high-quality twin for a large metropolitan area presupposes robust, high-capacity, low-latency communication links between sensors, platforms, and control systems.

Within this broader landscape, the physical communication layer of the smart-mobility infrastructure becomes a natural target for detailed modeling and optimization. The optical links that carry high-rate data between roadside equipment, aggregation points, and urban data centers act as arteries of the mobility nervous system. Their behavior under realistic loading, including the non-idealities of transceivers and optical amplifiers, constrains the kinds of AI-driven, real-time control loops that can be reliably deployed. The digital surrogate of an IM-DD optical communication system developed in this thesis is conceived as a building block within such mobility-oriented digital twins: a high-fidelity, hardware-aware model of one critical element of the digital infrastructure, designed to answer

⁴⁴ E. Ismagilova *et al.*, “Smart cities: Advances in research-An information systems perspective,” *International Journal of Information Management*, vol. 47, pp. 88-100, 2019; Y. Kaluarachchi, *op. cit.*

⁴⁵ Z. Wang *et al.*, “Mobility digital twin: Concept, architecture, case study, and future challenges,” *IEEE Internet of Things Journal*, vol. 9, pp. 17452-17467, 2022.

⁴⁶ F. Loia, C. Perillo, and G. Gravili, “Unleashing the power of digital twin and big data as a new frontier for smart mobility: An ecosystem perspective,” *Big Data Research*, 2025.

questions such as: *what latency and reliability margins can be expected under realistic traffic? How do hardware impairments impact link robustness? Moreover, how can signal-processing strategies, including AI-based equalization, be co-designed with the rest of the smart-city stack?*

2.3.4 Rome as a metropolitan example and the role of industrial actors (e.g. Almoviva)

Rome is a valuable case study for understanding how smart mobility strategies, digital infrastructures, and industrial actors converge in a large, complex European metropolis. Roma Capitale has adopted two complementary strategic frameworks: the *Roma Smart City* plan and the *Piano Urbano della Mobilità Sostenibile* (PUMS) at both city and metropolitan scale. The Roma Smart City plan is a programmatic document that provides a methodological and strategic framework for aligning priorities across multiple policy domains, including mobility, environment and waste, social services, culture, tourism, and safety, towards the 2030 vision for the city and its wider metropolitan area. It explicitly promotes a participatory, "living" planning instrument that integrates citizen needs and co-creation, rather than a purely top-down approach⁴⁷. The metropolitan PUMS further extends this perspective to the entire *Città metropolitana di Roma Capitale*, coordinating the core city with 120 surrounding municipalities in a unified strategy for environmentally, socially, and economically sustainable mobility⁴⁸.

Within this framework, Rome's sustainable mobility strategy combines structural changes in transport supply with a strong digital and governance component. The PUMS approved in 2022 sets explicit targets to increase the share of trips made with sustainable modes (public transport, cycling, walking, shared mobility) from roughly 37% to 52% by 2030. To achieve this, the plan foresees new tram and metro infrastructure, support for cycling and pedestrian mobility, expansion of shared mobility schemes, robust decarbonisation and electrification of

⁴⁷ Roma Capitale, "*Roma Smart City - Il Piano*," programmatic document of Roma Capitale, , available via the Comune di Roma website, 2021.

⁴⁸ Città metropolitana di Roma Capitale, "*Piano Urbano della Mobilità Sostenibile (PUMS) della Città metropolitana di Roma Capitale*," official PUMS portal, first approved 2023-2024.

both public and private fleets, and a systematic focus on road safety and accident reduction⁴⁹. These physical interventions are accompanied by investments in digitalization, "infomobility", and advanced traffic-management tools, so that citizens can plan and adapt their journeys using real-time information in a genuinely smart-city perspective⁴⁹.

European and national funding instruments (PON Metro, REACT-EU, RRF/PNRR, and others) have been used to upgrade Rome's mobility ecosystem, explicitly aligned with smart-city goals. Recent presentations by Roma Servizi per la Mobilità highlight the development of new low-emission zones, traffic-monitoring systems, a modernized public transport fleet, an upgraded Mobility Control Room, new data-management applications and predictive systems, all conceived as building blocks for forthcoming Smart Road and Smart City projects and for the diffusion of Mobility-as-a-Service (MaaS) solutions at the national level⁵⁰. At the same time, participation in European initiatives such as the UPPER project (Mission Cities, aiming for climate neutrality by 2030) frames Rome's mobility policies within broader EU climate and innovation agendas, reinforcing the link between sustainable urban mobility, digital transition, and long-term decarbonization commitments⁵¹.

In this context, industrial players in the ICT and mobility sectors, such as Almoviva S.p.A., play a crucial enabling role. Almoviva positions itself as a leading Italian Information & Communication Technology group, with a dedicated *Citizenship & Government* business line focused on building sustainable, interconnected, and inclusive communities by supporting the digital transformation of public administrations and territories. The group reports long-standing

⁴⁹ A. Donati, "Roma Capitale: Smart Mobility per la mobilità sostenibile," Energia, Ambiente e Innovazione (ENEA), overview of the PUMS and Rome's sustainable mobility strategy to 2030, 2022.

⁵⁰ F. Nussio, "Development of mobility in Rome with REACT-EU structural funds," presentation for Barcelona Smart City Expo, 2023, PON Metro / REACT-EU / RRF funding overview and Control Room development for Roma Capitale.

⁵¹ Roma Capitale, "Verso un futuro a zero emissioni - Il Progetto UPPER e le sue misure," institutional magazine, , outlining Rome's role in the UPPER Mission Cities project and its climate-neutrality objectives for 2030, 2025.

experience in urban digitalization, with hundreds of active clients and large-scale deployments spanning digital public services, mobility, cultural heritage, and environmental applications⁵².

More specifically on mobility, Almoviva has developed *MOOVA*, a technologically advanced platform for sustainable, multimodal, "as-a-service" mobility. MOOVA is designed to manage complex mobility ecosystems with stringent requirements on operational efficiency, interoperability among multiple stakeholders, and high service-quality expectations⁵³. Within MOOVA, solutions such as *MOOVA Passenger Flow* integrate heterogeneous data sources (on-board sensors, roadside systems, external providers like telecom operators) to monitor passenger numbers and crowding in near real time, and to generate predictive analytics that allow operators to reschedule services and inform users dynamically⁵⁴. These capabilities directly mirror the data-intensive, AI-enabled, and citizen-centric requirements discussed in the previous section: they depend on robust communication networks, scalable back-end processing, and advanced analytics to translate raw data into operational decisions.

Rome itself has become a showcase for this type of collaboration between public institutions and ICT providers. At the Smart City Expo World Congress in Barcelona, Roma Capitale has presented a new Command & Control Room for traffic management, developed by Almoviva on top of the MOOVA platform for Roma Servizi per la Mobilità. This platform integrates multiple applications and data streams to monitor traffic in real time, support sustainable transport development, and improve the quality of urban mobility services⁵⁵. The project is frequently presented as an example of "data-driven Rome", where real-time

⁵² Almoviva S.p.A., "Citizenship & Government - Urban Future," corporate website, accessed; overview of Almoviva's activities in smart cities, smart mobility and digital public services, 2025.

⁵³ Almoviva S.p.A., "MOOVA: Almoviva's platform for sustainable, multimodal, as-a-service mobility," in TTS Italia Newsletter, 28 June 2022.

⁵⁴ Almoviva S.p.A., "Play it safe with MOOVA Passenger Flow," TTS Italia Newsletter, 28 April 2022

⁵⁵ Almoviva S.p.A., "The Development of Sustainable Transportation in the Cities of the Future - Almoviva with Roma Capitale at the Smart City Expo World Congress in Barcelona," press release, 6 November 2024.

information, predictive tools, and integrated control are used to manage congestion, support public transport, and prepare for future Smart Road and MaaS deployments⁵⁶.

From this thesis's perspective, Rome's trajectory and Almoviva's role illustrate how smart-city and smart-mobility ambitions ultimately rest on the reliability and performance of the underlying digital and communication infrastructures. Large-scale platforms such as MOOVA and Rome's traffic Command & Control Room are only as robust as the networks and data centers that connect sensors, vehicles, roadside units, and control applications. High-capacity, low-latency communication links, such as short-reach optical interconnects in data centers, control rooms, and core urban networks, are a critical yet often unseen part of this infrastructure. The digital surrogate developed in this work can be interpreted as a contribution at this "deep infrastructure" level: by providing a high-fidelity, hardware-aware model of IM-DD optical links, it supports the design and optimization of the physical communication layer on which smart-city traffic management platforms, such as those deployed in Rome, ultimately depend.

2.4 Optical Communication as Enabler

2.4.1 ICT backbone: from sensors to data centers

Smart urban mobility and traffic management depend on a multi-layer ICT backbone that connects sensors, roadside units, vehicles, control centers, and data centers. At the edge, large numbers of heterogeneous devices, such as fixed sensors, roadside units, connected vehicles, and public-transport fleets, generate high-rate data streams that must be transmitted reliably and with low latency⁵⁷. For applications such as real-time traffic control, cooperative awareness, and safety-critical vehicle-to-everything (V2X) services, the communication requirements are

⁵⁶ Roma Servizi per la Mobilità & Almoviva, "Roma Data Driven" workshop/video session, Smart City-related events on data-driven mobility in Rome, 2024–2025.

⁵⁷ D. Oladimeji *et al.*, "Smart Transportation: An Overview of Technologies and Applications," *Sensors*, 23, 2023.

stringent: end-to-end delays on the order of milliseconds, high reliability, and sufficient bandwidth to carry dense sensor and video traffic⁵⁸.

These requirements have led to an architecture in which wireless access (cellular, Wi-Fi, V2X) is tightly coupled to a high-capacity optical backbone. Wireless links provide flexible, ubiquitous connectivity to mobile users and vehicles, while optical fibre networks aggregate and transport the resulting traffic across the metropolitan area. Systematic studies of smart-city and 5G deployments emphasize that fibre-based fronthaul, backhaul, and core networks are essential to support network slicing, edge computing, and ultra-reliable low-latency communication for ITS services⁵⁹. In practice, this means that traffic data collected at intersections and along road corridors is forwarded over short and medium-reach optical links towards edge nodes and data centers, where analytics and decision logic are executed.

A second trend is the integration of vehicular edge computing with this optical backbone. Rather than sending all data to a distant cloud, part of the processing is offloaded to edge servers located near the road network or integrated into roadside units⁶⁰. Surveys on the Internet of Vehicles and vehicular edge computing show architectures in which vehicles, access points, and edge servers are tightly coupled, with fibre-based aggregation networks providing the low-latency interconnection between edge clusters and central control platforms⁶¹. From the perspective of this thesis, the optical backbones that connect the sensing layer to edge and core data center resources are critical areas in which short-reach IM/DD optical links play a significant enabling role. Consequently, these links provide an ideal environment for developing the digital surrogate in the following chapters.

⁵⁸ B. Ji *et al.*, “Survey on the Internet of Vehicles: Network Architectures and Applications,” *IEEE Communications Standards Magazine*, 4, 34–41, 2020.

⁵⁹ C. Yang *et al.*, “Using 5G in Smart Cities: A Systematic Mapping Study,” *Intelligent Systems with Applications*, 14, 200065, 2022

⁶⁰ L. Liu *et al.*, “Vehicular Edge Computing and Networking: A Survey,” *Mobile Networks and Applications*, 26, 1145–1168, 2021

⁶¹ *Ibidem*

2.4.2 Short-reach IM/DD optical links in data centers and edge clouds

The logical core of modern traffic-management and smart-mobility platforms resides in data centers and edge clouds, which must ingest and process large volumes of streaming data from the road network. Short-reach optical links, typically using intensity-modulated direct-detection (IM/DD) technologies, serve as the workhorses in these environments, offering high throughput over distances ranging from a few meters to several kilometers at relatively low cost and power consumption⁶². Recent reviews of short-reach optical communications emphasize that IM/DD links, together with emerging coherent approaches, underlie the dense interconnect fabric of data centers and edge sites that support latency-sensitive applications⁶².

For smart cities, the edge-computing layer is critical. Surveys on edge-enabled smart cities describe architectures in which local or regional edge nodes perform real-time analytics on traffic, video, and sensor data, offloading only aggregated information or model updates to central clouds⁶³. In these designs, short-reach optical interconnects connect servers within the edge cluster and link the cluster to the metro or core network, ensuring that analytics and control loops for traffic signals, incident detection, and V2X coordination can operate within tight latency budgets⁶³. From the perspective of ITS, this means that the performance and robustness of IM/DD links inside data centers and between edge sites directly impact the responsiveness of traffic-management services.

On the access side, time-sensitive networking (TSN) and 5G/V2X deployments rely on optical fronthaul and backhaul to connect base stations and roadside units to edge and core clouds. A recent systematic review of TSN and 5G integration highlights the importance of deterministic, low-jitter transport between radio access nodes and computing resources, often implemented via fibre-based segments⁶⁴. In

⁶² A. Karar, A. Falou, J. Barakat, Z. Gürkan, and K. Zhong, “Recent Advances in Coherent Optical Communications for Short-Reach: Phase Retrieval Methods,” *Photonics*, 10(3), 308, 2023.

⁶³ L. Khan *et al.*, “Edge-Computing-Enabled Smart Cities: A Comprehensive Survey,” *IEEE Internet of Things Journal*, 7, 10200–10232, 2020.

⁶⁴ Z. Satka *et al.*, “A comprehensive systematic review of integration of time sensitive networking and 5G communication,” *Journal of Systems Architecture*, 138, 102852, 2023.

combination with 5G-enabled vehicle-to-everything (V2X) systems for smart cities, these short- and medium-range optical links create a reliable path for transporting aggregated vehicular and traffic data to the central decision-making cores⁶⁴. The digital surrogate developed in this thesis targets IM/DD-based short-reach links, focusing on capturing device non-idealities and noise sources that might otherwise be abstracted away in high-level smart city models.

2.4.3 Energy efficiency, resilience, and circular-economy implications of optical networks

Optical communication networks are increasingly regarded as essential to the development of sustainable, energy-efficient digital infrastructures, in addition to their performance advantages. Reviews on 5G and smart cities consistently report that fibre-based backbones offer significantly higher energy efficiency than copper or purely wireless alternatives, owing to lower transmission losses and much higher capacity per unit of power⁶⁵. In integrated 5G/ITS architectures, optical backhaul and core networks therefore play a central role in containing the overall energy consumption of the connectivity layer while still supporting high data rates and low latency⁶⁵.

This has significant implications for urban transportation and mobility systems. As traffic management, public transport, and V2X services become more data-intensive, the energy consumption of digital infrastructure becomes part of the urban sustainability equation. Studies on the energy use of 5G networks show that, while radio access can be energy-hungry, careful end-to-end network design, including efficient optical transport, can substantially mitigate the overall carbon footprint of digital services⁶⁶. In the vehicular domain, analyses of the “Green Internet of Vehicles” show that communication and computing architectures optimized for energy efficiency, often leveraging fibre-based aggregation networks,

⁶⁵ M. Shehab *et al.*, “5G Networks Towards Smart and Sustainable Cities: A Review of Recent Developments, Applications and Future Perspectives,” *IEEE Access*, 10, 2987–3006, 2022.

⁶⁶ L. Williams, B. Sovacool, and T. Foxon, “The energy use implications of 5G: Reviewing whole network operational energy, embodied energy, and indirect effects,” *Renewable and Sustainable Energy Reviews*, 157, 112033, 2022.

are essential to ensure that connected-vehicle services themselves do not undermine transport decarbonisation goals⁶⁷.

From a circular-economy perspective, energy-efficient optical networks reduce operational energy demand and enable more resource-efficient urban systems. Technical reviews on smart-city energy management and the Internet of Energy argue that low-loss, high-capacity communication links are necessary to coordinate distributed energy resources, demand response, and electrified transport fleets⁶⁸. In parallel, scoping reviews on 5G and climate policy discuss how digital networks can support climate change mitigation, provided that their life-cycle impacts are managed and that they facilitate significant emissions reductions in other sectors, such as transportation⁶⁹. Optical ICT backbones are thus doubly relevant: they reduce the footprint of digital infrastructure and enable advanced mobility and energy management services that facilitate circular, low-carbon urban systems.

This thesis focuses on energy-efficient IM/DD optical links, aligned with this broader agenda. By building a high-fidelity digital surrogate of short-reach optical interconnects, we aim not only to optimize performance and robustness for smart-mobility use cases but also to provide a tool that can later be used to explore energy-performance trade-offs and inform greener design choices at the physical-layer level.

2.5 The digital surrogate in this ecosystem

2.5.1 Role of the digital surrogate for an IM-DD BPAM link

The previous sections have outlined how smart-city mobility platforms depend on dense digital infrastructures: roadside and in-vehicle sensors, 5G/6G access, edge computing nodes, and data centers that host decision-making and optimization algorithms. In all of these environments, a significant fraction of traffic is carried

⁶⁷ J. Wang, K. Zhu, and E. Hossain, “Green Internet of Vehicles (IoV) in the 6G Era: Toward Sustainable Vehicular Communications and Networking,” *IEEE Transactions on Green Communications and Networking*, 6, 391–423, 2022.

⁶⁸ P. Mishra and G. Singh, “Energy Management Systems in Sustainable Smart Cities Based on the Internet of Energy: A Technical Review,” *Energies*, 16(19), 6903, 2023.

⁶⁹ G. Huseien and K. Shah, “Potential Applications of 5G Network Technology for Climate Change Control: A Scoping Review of Singapore,” *Sustainability*, 13(17), 9720, 2021.

over short-reach intensity-modulated direct-detection (IM-DD) optical links that connect switches, servers, and accelerators within data centers and edge clouds, and that provide the optical backhaul and fronthaul for radio and vehicular networks⁷⁰. These links must simultaneously satisfy stringent latency, throughput, and reliability targets while respecting tight constraints on cost and energy per bit⁷¹.

In this thesis, the focus is narrowed to a single representative class of such links: a high-speed IM-DD optical channel employing bipolar multilevel pulse-amplitude modulation (BPAM) and operated with realistic transmitter, channel, and receiver hardware. Rather than treating each hardware block (DAC, driver, MZM, fiber, SOA preamplifier, photodiode/TIA, ADC, DSP) in isolation, we seek to reproduce the end-to-end behavior of the entire chain, including the cumulative effects of nonidealities, noise, and nonlinearities. The resulting model is referred to as a digital surrogate: a modular, parameterized numerical replica of the specific laboratory link, designed to behave like the physical system under a wide range of operating conditions.

This concept is related to, but more focused than, the broader notion of a digital twin for optical networks. Recent work on digital twins in optical communication typically focuses on network- or system-level constructs that mirror an entire optical transport network, often with the goal of real-time monitoring, fault management, or intent-based control⁷². In contrast, the digital surrogate developed in this thesis targets a single IM-DD link. It provides high physical fidelity at the block level (including DAC/ADC nonidealities and SOA dynamics). It is primarily used as an offline design, analysis, and training environment rather than as a continuously synchronized operational twin⁷³. This narrower scope is deliberate: it

⁷⁰ C. Zhang *et al.*, “Recent Advances of High-Speed Short-Reach Optical Interconnects for Data Centers,” IEEE Open Journal of the Solid-State Circuits Society, 5, 86-100.

⁷¹ C. Pang *et al.*, “200-Gb/s-Lane IM-DD Technologies for Short-Reach Optical Interconnects,” Journal of Lightwave Technology, vol. 38, no. 2, pp. 492-503, 2020.

⁷² Q. Zhuge *et al.*, “Building a Digital Twin for Intelligent Optical Networks [Invited Tutorial],” Journal of Optical Communications and Networking, vol. 15, no. 8, pp. C242-C262, 2023.

⁷³ S. Wang *et al.*, “Role of Digital Twin in Optical Communication: Fault Management, Hardware Configuration, and Transmission Simulation,” IEEE Communications Magazine, vol. 59, no. 1, pp. 108-114, 2021.

matches the scale at which most design decisions for short-reach optical links are made (modulation format, baud rate, equalization strategy, operating point of the SOA, DAC/ADC resolution, and so on), while remaining consistent with the general digital-twin paradigm explored in the optical-network literature⁷².

2.5.2 Optimized BPAM links as enablers of sustainable smart-city infrastructures

From the perspective of Industry 5.0 and the circular economy, the key contribution of this work is not the digital surrogate *per se*, but the class of optical solutions it enables and de-risks. Modern data-center and edge-cloud interconnects increasingly rely on multilevel PAM formats (e.g., PAM4 and higher-order PAM) to push per-lane data rates beyond 100-200 Gb/s while keeping transceiver complexity and cost compatible with large-scale deployment⁷⁴. Advanced IM-DD schemes, combined with carefully optimized transmitter and receiver front-ends, allow operators to extract more throughput from the same optical infrastructure, reduce the number of parallel fibers and transceivers, and thereby lower both capital and operational expenditures⁷⁵.

In this thesis, a bipolar PAM variant (BPAM) is investigated to further improve the trade-off among spectral efficiency, noise tolerance, and implementation complexity. Although BPAM is not yet a mainstream industrial standard, it belongs to the broader family of multilevel IM-DD formats that have been studied for data-center and short-reach links. The digital surrogate is the tool that allows us to explore this design space systematically and safely: it provides a controlled environment to assess how BPAM behaves under realistic DAC/ADC limitations, SOA saturation, and receiver noise, and to co-design classical and ML-based DSP for equalization and pre-distortion.

When such link-level optimizations are implemented in deployed systems, their implications extend to the macro-level goals discussed earlier. Energy-efficient,

⁷⁴ C. Pang *et al.*, “200-Gb/s-Lane IM-DD Technologies for Short-Reach Optical Interconnects,” *Journal of Lightwave Technologies*, vol. 38, no. 2, pp. 492-503, 2020.

⁷⁵ C. Zhang *et al.*, “Recent Advances of High-Speed Short-Reach Optical Interconnects for Data Centers,” *IEEE Open Journal of the Solid-State Circuits Society*, 5, 86-100, 2025.

high-capacity optical links reduce power consumption and equipment footprint in data centers and edge nodes that host smart-city and ITS services⁷⁶. Reviews of 5G and vehicular communications consistently identify optical backbones as a key enabler for reducing the energy per transported bit and for supporting “green” Internet-of-Vehicles and smart-city architectures⁷⁷. From a circular-economy perspective, higher spectral efficiency and improved OSNR budgets can extend the useful life of installed fiber and transceiver hardware, reduce the need for overdimensioning, and facilitate incremental upgrades rather than complete equipment replacement. The digital surrogate enables us to design and validate such link configurations before committing to hardware, thereby reducing both experimental overhead and material waste.

2.5.3 Indirect link to smart-city mobility and traffic management

The connection between this link-level work and smart-city mobility is therefore indirect but structurally significant. As discussed in Sections 2.3 and 2.4, urban traffic management and innovative mobility platforms rely on a multi-tier digital architecture: sensors and vehicles at the edge; 5G/6G and Wi-Fi access networks; optical fronthaul and backhaul; and a hierarchy of edge and core data centers running AI-based analytics and control algorithms⁷⁸. The IM-DD BPAM link modeled in this thesis represents one instantiation of the optical “pipeline” within these data center and edge cloud facilities, particularly for short-reach interconnects that must sustain high traffic loads from ITS and mobility applications while preserving low latency and high availability.

By providing a realistic, end-to-end surrogate of such a link, the thesis offers a design and experimentation platform for evaluating new modulation formats, DSP

⁷⁶ M. J. Shehab *et al.*, “5G Networks Towards Smart and Sustainable Cities: A Review of Recent Developments, Applications and Future Perspectives,” *IEEE Access*, vol. 10, pp. 2987-3006, 2022.

⁷⁷ J. Wang, K. Zhu, and E. Hossain, “Green Internet of Vehicles (IoV) in the 6G Era: Toward Sustainable Vehicular Communications and Networking,” *IEEE Transactions on Green Communications and Networking*, vol. 6, no. 1, pp. 391-423, 2022.

⁷⁸ L. U. Khan *et al.*, “Edge-Computing-Enabled Smart Cities: A Comprehensive Survey,” *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 10200-10232, 2020.

algorithms, and operating points under conditions that match laboratory hardware and, by extension, the types of transceivers used in smart-city infrastructures. The subsequent chapters zoom in on the physics, modeling, and optimization of this BPAM IM-DD link. The broader motivations behind smart cities and Industry 5.0, which were discussed in this chapter, provide the context for interpreting the technical results. The goal is to develop a more capable, efficient, and robust optical building block that will strengthen the digital infrastructure supporting sustainable and resilient urban mobility services.

Chapter 3

Background & State of the Art

This chapter provides the technical background and state-of-the-art context for the digital surrogate developed in this thesis. Building on the broader framework of Industry 5.0, circular economy, and smart-city mobility introduced in Chapter 2, the focus now narrows to the physical and algorithmic foundations of high-speed intensity-modulated direct-detection (IM-DD) optical links. The goal is to explain the construction of modern short-reach systems, the impairments that affect their performance, the common modeling approaches used for these systems, and the application of DSP and machine-learning techniques to mitigate these limitations.

The chapter is structured to follow the signal flow of a generic short-reach IM-DD link. Section 3.1 reviews the main building blocks, from digital pre-processing through DAC, driver, modulator, fiber, and SOA, to PD/TIA, ADC, and receiver DSP, and introduces key performance metrics such as BER, OSNR, eye diagrams, and EVM. Section 3.2 discusses the most relevant device-level non-idealities and impairments in these blocks, with emphasis on DAC/ADC quantization and nonlinearities, bandwidth constraints, SOA saturation and ASE noise, and receiver noise sources. Section 3.3 surveys existing modeling and simulation approaches, including commercial link simulators, analytical models, and vendor tools, and highlights the simplifications they typically adopt. Section 3.4 reviews the emerging literature on digital twins and digital surrogates in communications and photonics, clarifying their interpretation at the network and system levels. Section 3.5 summarizes the state of the art in receiver DSP and machine-learning-based processing for IM-DD links, ranging from classical linear and nonlinear equalization to neural network-based equalizers and pre-distortion schemes. Finally, Section 3.6 synthesizes these strands into a gap analysis that motivates the thesis's specific contribution.

Throughout this chapter, particular emphasis is placed on multilevel pulse-amplitude modulation, especially bipolar PAM (BPAM), which serves as the primary modulation format examined in the subsequent sections of the work.

BPAM is treated here as part of the broader family of multilevel IM-DD formats deployed or proposed for short-reach applications, and its positioning with respect to conventional unipolar PAM4/PAM8 is discussed where relevant. This establishes the essential conceptual and bibliographic groundwork for the detailed modeling, digital surrogate design, and experimental studies presented in the following chapters.

3.1 IM-DD optical links: architecture and performance metrics

3.1.1 System architecture of short-reach IM-DD optical links

Short-reach intensity-modulated direct-detection (IM-DD) links are the workhorse of contemporary intra-data-center and short-reach metro/access connections. They are preferred wherever cost, power consumption, and form-factor constraints make coherent detection unattractive. At the same time, the required line rates and reach remain compatible with intensity modulation and square-law detection¹. In this regime, IM-DD links typically operate at per-lane gross bit rates from tens up to a few hundred gigabits per second, using multilevel pulse-amplitude modulation (PAM) formats such as PAM4 or higher-order variants².

A generic short-reach IM-DD link can be viewed as a cascade of functional blocks that map a digital bitstream into an optical intensity waveform, transport it over fiber, and recover the information at the receiver. Figure 3.1 will illustrate the reference architecture adopted in this thesis:

¹ Y. Xie, Y. Wang, S. Kandeepan, & K. Wang, "Machine learning applications for short reach optical fiber communications", *Photonics*, vol. 9, no. 1, p. 30, 2022.

² X. Pang, *et al.* "200 Gbps/lane IM/DD technologies for short reach optical interconnects", *Journal of Lightwave Technology*, vol. 38, no. 2, pp. 492–503, 2020.

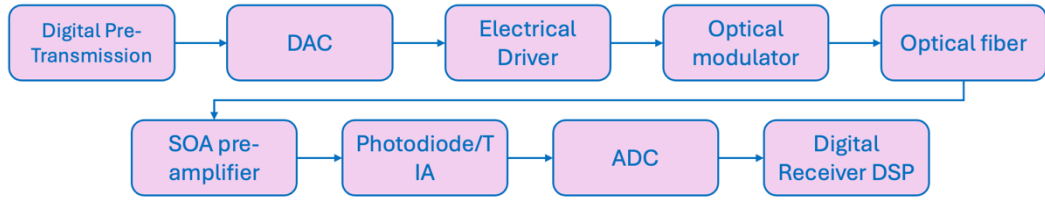


Figure 3.1 Reference architecture

On the transmitter side, digital pre-transmission processing starts from a binary bitstream and produces a sequence of symbols suitable for IM-DD transmission. This stage typically includes bit-to-symbol mapping (e.g., to PAM4, PAM8, or bipolar PAM constellations), optional framing and FEC encoding, digital pre-emphasis or equalization to partially compensate for bandwidth limitations, and possibly digital predistortion to mitigate known nonlinearities in the electrical and optical front end³. The output of this stage is a discrete-time waveform in the digital domain.

The digital-to-analog converter (DAC) transforms the discrete-time digital waveform into an analog electrical signal at a high sampling rate. Practical DACs introduce finite resolution (quantization), bandwidth limitations, and dynamic-range constraints, all of which directly affect the achievable modulation fidelity at high baud rates. The DAC output is then fed into an electrical driver, which amplifies and conditions the signal to the voltage levels and impedance required by the optical modulator, while minimizing distortion and noise.

The optical modulator converts the electrical drive signal into an intensity-modulated optical waveform. Depending on the implementation, this can be a Mach-Zehnder modulator (MZM), an electro-absorption modulated laser (EML), or a directly modulated laser (DML). In this thesis, the reference system is based on an external MZM driven around an appropriate bias point to implement a practical IM-DD transfer function. The modulator shapes the optical carrier emitted by a continuous-wave (CW) laser into the desired intensity pattern, thereby mapping electrical PAM (or BPAM) symbols to optical power.

³ *Ibidem*.

The modulated signal propagates along an optical fiber segment whose length is typically from a few hundred meters to a few tens of kilometers, depending on the application (intra-data center, data center interconnect, access, or metro). Even over relatively short distances, chromatic dispersion, attenuation, and other linear effects impact the waveform and must be accounted for in the link design. In some scenarios, especially when link budgets are tight or higher-loss components are present, an optical semiconductor optical amplifier (SOA) is inserted as a preamplifier before the receiver. The SOA boosts the received power but introduces gain saturation, pattern-dependent distortion, and amplified spontaneous emission (ASE) noise, whose effects must be carefully modeled in high-speed IM-DD links.

At the receiver, a photodiode and transimpedance amplifier (PD/TIA) perform square-law detection, converting optical intensity into an electrical current, which is then converted into a voltage. This stage is subject to shot noise, thermal noise, relative intensity noise (RIN), and bandwidth limitations that jointly shape the received electrical signal. The analog waveform is then sampled and quantized by an analog-to-digital converter (ADC), which, like the DAC, has finite resolution, input range, and bandwidth. Finally, the digital receiver DSP performs clock and data recovery, equalization (e.g., linear or nonlinear equalizers), symbol decision, and error counting or decoding. In more advanced designs, this block can also host neural-network-based equalizers or other machine-learning components, as discussed later in the chapter.

From the viewpoint of this thesis, this end-to-end architecture serves two purposes. First, it defines the physical signal path that the digital surrogate must emulate, with clear domain transitions (digital $\rightarrow$ electrical $\rightarrow$ optical $\rightarrow$ electrical $\rightarrow$ digital) and well-defined interfaces between blocks. Second, it provides a natural framework for positioning different modulation formats, including conventional unipolar PAM4/PAM8 and emerging bipolar PAM (BPAM) schemes, which can be viewed as alternative symbol mappings and waveform shapes injected at the digital pre-transmission stage and propagated through the same chain of components.

3.1.2 Modulation formats and performance metrics in IM-DD links

In short-reach optical interconnects, intensity modulation with direct detection has converged on multilevel pulse-amplitude modulation (PAM) as the dominant format family for scaling line rates under tight cost and complexity constraints⁴. In its simplest form, binary PAM (2-PAM) corresponds to the classic non-return-to-zero (NRZ) on-off keying used in legacy systems, while higher-order formats such as PAM4, PAM6, and PAM8 encode multiple bits per symbol by mapping bit patterns onto a finite set of amplitude levels. For instance, PAM4 encodes two bits per symbol using four intensity levels, effectively doubling the bit rate at a fixed symbol rate, whereas PAM8 encodes three bits per symbol using eight levels, and so on.

Mathematically, an M-ary PAM signal can be described as a sequence of real-valued symbols $\{a_k\}$ drawn from a discrete alphabet $\mathcal{A} = \{a_0, \dots, a_{M-1}\}$, modulating a base pulse shape $p(t)$ at symbol period T_s :

$$x(t) = \sum_k a_k p(t - kT_s).$$

For unipolar IM-DD systems, the symbols are typically chosen as non-negative levels, e.g., $a_m = m\Delta$ with $m = 0, \dots, M - 1$ and step size Δ , to ensure that the optical intensity at the modulator output remains positive. The average optical power and the minimum Euclidean distance between levels (which controls noise tolerance) are constrained by the link budget and hardware. As M increases, the spectral efficiency (bits per symbol) improves. Still, the minimum distance between adjacent levels shrinks, making the format more sensitive to noise, non-linearities, bandwidth limitations, and quantization.

In current Ethernet and data-center standards, PAM4 has become the baseline modulation for 50–200 Gb/s per-lane IM-DD links, with higher-order PAM (e.g., PAM6/PAM8 with probabilistic or geometric shaping) under active investigation to reach 200–400 Gb/s per lane⁵. These formats must operate within the limited

⁴ K. Zhong, X. Zhou, J. Huo, C. Yu, C. Lu, A. P. T. Lau, “Digital Signal Processing for Short-Reach Optical Communications: A Review of Current Technologies and Future Trends”, *Journal of Lightwave Technology*, vol. 36, no. 2, pp. 377–400, 2018.

⁵ *Ibidem*.

electrical bandwidth of DACs, drivers, modulators, photodiodes, and ADCs, which imposes a practical ceiling on the symbol rate and therefore motivates the use of multilevel constellations rather than simply increasing baud rate.

A more recent development, directly relevant to this thesis, is optical bipolar PAM (BPAM). In BPAM, the symbol alphabet is symmetric around zero, e.g., $\mathcal{A} = -3, -1, +1, +3$ for BPAM-4, so that information is encoded both in amplitude and in the *sign* of the electrical waveform. At first glance, such bipolar signals are incompatible with square-law intensity detection, which nominally discards phase information and removes sign information. However, theoretical⁶ and experimental⁷ work has shown that, with oversampling by a factor of 2, it is possible to reconstruct the sign of the underlying field and thus recover a bipolar constellation using a single photodiode and a standard IM-DD front-end, without coherent detection. In this way, BPAM effectively “re-introduces” a phase degree of freedom into IM-DD links, enabling constellations that are closer in spirit to coherent schemes but with a much simpler optical receiver.

From a systems perspective, BPAM and shaped PAM formats can be interpreted as different symbol mappings on the same IM-DD link architecture introduced in Section 3.1.1. The physical chain (DAC $\rightarrow$ driver $\rightarrow$ modulator $\rightarrow$ fiber $\rightarrow$ optional SOA $\rightarrow$ PD/TIA $\rightarrow$ ADC $\rightarrow$ DSP) remains unchanged; only the statistics and geometry of the transmitted symbol sequence $\{a_k\}$ are modified. This viewpoint is particularly convenient for a digital surrogate: once the blocks and interfaces are defined, alternative constellations and pulse shapes can be studied as “*plug-in*” options at the digital transmitter, with their impact propagated through the same non-ideal components.

To compare such formats quantitatively, several performance metrics are routinely used:

⁶ M. Secondini and E. Forestieri, “Direct Detection of Bipolar Pulse Amplitude Modulation”, *Journal of Lightwave Technology*, vol. 38, no. 21, pp. 5981–5990, 2020.

⁷ M. Secondini, E. Forestieri, F. Fresi, L. Poti, T. Catuogno, and F. Cavaliere, “Bipolar Pulse Amplitude Modulation with Direct Detection”, in *Proc. European Conference on Optical Communication (ECOC)*, paper cp.2019.0926, 2019.

- **Bit rate and symbol rate.** If an M-PAM constellation carries $\log_2 M$ bits per symbol at symbol rate $R_s = 1/T_s(\text{baud})$, the gross line rate is $R_b = R_s \log_2 M$. For fixed hardware bandwidth, moving from PAM4 to PAM8 or BPAM formats increases R_b without increasing R_s , at the price of tighter SNR requirements.

- **Signal-to-noise ratio (SNR) and optical signal-to-noise ratio (OSNR).** In amplified links, performance is often expressed as OSNR, defined as the ratio of signal power to amplified spontaneous emission (ASE) noise power within a reference bandwidth (e.g., 0.1 nm). In short-reach IM-DD links without inline amplification, the electrical SNR after detection and front-end filtering is more commonly used. Analytical and experimental studies show that each increment in PAM order typically incurs several dB of SNR/OSNR penalty at a given bit-error rate (BER), due to the reduced eye opening between adjacent levels⁸.

- **Bit-error rate (BER) and symbol-error rate (SER).** BER (or SER) as a function of OSNR/SNR is the primary metric for comparing modulation formats. In practical systems, one often considers the *pre-FEC* BER, i.e., the error rate measured before forward-error-correction (FEC) decoding, which must lie below the threshold of the employed forward-error-correction (FEC) code (e.g., around 10^{-3} to 10^{-2} depending on the code)⁹. [FN4] A detailed treatment of BER/SER for PAM and other digital modulation formats can be found in standard digital communication textbooks¹⁰.

- **Error-vector magnitude (EVM).** Especially in short-reach standards, EVM is widely used as a compact measure of signal quality. It quantifies the root-mean-square difference between received constellation points and their ideal positions, normalized by the average symbol energy.

⁸ K. Zhong, X. Zhou, J. Huo, C. Yu, C. Lu, A. P. T. Lau, “Digital Signal Processing for Short-Reach Optical Communications: A Review of Current Technologies and Future Trends”, *Journal of Lightwave Technology*, vol. 36, no. 2, pp. 377–400, 2018.

⁹ J. G. Proakis and M. Salehi, “Digital Communications”, 5th ed., McGraw-Hill, 2008.

¹⁰ *Ibidem*

Tutorials on DSP for short-reach IM-DD links show how EVM correlates with BER and how it can be decomposed into contributions from noise, bandwidth limitation, non-linearities, and quantization¹¹.

- **Eye-diagram metrics.** Time-domain eye diagrams remain a central tool for IM-DD links. Metrics such as eye height, eye width, crossing level, and jitter provide an intuitive visualization of intersymbol interference, noise, and timing impairments. For multilevel and bipolar formats, multi-level eyes or reconstructed “virtual” eyes at the symbol decision point are often used to diagnose implementation penalties¹².

In summary, short-reach IM-DD systems form a rich design space where modulation format, symbol rate, pulse shape, and DSP are co-optimized under stringent hardware constraints. Conventional unipolar PAM4/PAM8 formats push the limits of SNR and bandwidth, while emerging schemes such as BPAM aim to exploit additional degrees of freedom (e.g., implicit phase) within the IM-DD paradigm. The digital surrogate developed in this thesis is designed to capture relevant performance metrics, BER, EVM, eye-diagram figures, and OSNR/SNR, for both conventional and bipolar formats, enabling controlled, physically meaningful exploration of these trade-offs.

3.2 Device non-idealities and impairments

3.2.1 DAC non-idealities: quantization, linearity, dynamic errors, and noise

In a short-reach IM-DD transmitter, the digital-to-analog converter (DAC) is the first active block, converting digital symbols into an analog electrical waveform. Its effective resolution, bandwidth, and non-idealities set a hard limit on how cleanly multilevel PAM and BPAM signals can be generated before pre-emphasis, driving, and optical modulation. Modern links targeting 50-200 Gbaud symbol rates typically employ high-speed current-steering or segmented DAC architectures,

¹¹ *Ibidem*.

¹² G. P. Agrawal, “*Fiber-Optic Communication Systems*”, 5th ed., Wiley, 2021.

which are optimized for bandwidth but exhibit a range of static and dynamic errors¹³.

From a modeling perspective, it is convenient to distinguish five main impairment classes:

- Quantization (finite resolution)
- Static non-linearities (INL/DNL)
- Dynamic errors and glitches
- Thermal and reference noise
- Timing jitter

Together, these effects determine the effective number of bits (ENOB) of the DAC and introduce non-Gaussian distortions that are not captured by a simple “AWGN + ideal quantizer” model¹⁴.

3.2.1.1 Quantization

In an ideal N-bit DAC, the input code $k \in 0, \dots, 2^N - 1$ is mapped to an output voltage (or current) level.

$$v_k = V_{\min} + k\Delta, \quad \Delta = \frac{V_{\max} - V_{\min}}{2^N - 1},$$

With a uniform step size Δ . For a full-scale signal, the quantization error, which is the difference between the original continuous analog signal value and its closest discrete digital representation, can be modeled, under standard assumptions, as a white random variable uniformly distributed in $[-\Delta/2, \Delta/2]$, whose variance is $\sigma_q^2 = \Delta^2/12$ ¹⁵. [FN2] For a sinusoidal test signal, this leads to the familiar approximation of signal-to-quantization-noise ratio

$$\text{SNR}_q \approx 6.02, N + 1.76 \text{ dB},$$

which provides a useful first-order benchmark for ENOB and dynamic range.

¹³ X. Li & L. Zhou, “A survey of high-speed and high-resolution current-steering DACs,” *Journal of Semiconductors*, vol. 41, no. 11, art. 111404, 2020.

¹⁴ *Ibidem*.

¹⁵ L. Michaeli and J. Šaliga, “Error models of the analog to digital converters,” *Measurement Science Review*, vol. 14, no. 2, pp. 62-77, 2014.

In multilevel PAM and BPAM waveforms, the situation is more delicate. The symbol alphabet occupies only part of the full-scale range; pre-emphasis and pulse-shaping filters can produce peaks beyond the nominal levels, and the signal is far from sinusoidal. As a result, the effective quantization noise depends on how the discrete codewords are visited over time, and on how closely the constellation levels align with DAC codes. For higher-order PAM and BPAM formats, small quantization steps are critical for preserving eye openings between densely packed levels, especially after linear pre-emphasis has boosted high-frequency components.

3.2.1.2 Static linearity: INL and DNL

Real DACs deviate from the ideal transfer characteristic due to mismatches among current sources, resistor segments, and switching devices. These static imperfections are typically characterized by *differential nonlinearity (DNL)* and *integral nonlinearity (INL)*. DNL at code k measures the deviation of the actual step size ($\Delta_k = v_k - v_{k-1}$) from the ideal least significant bit (LSB):

$$\text{DNL}(k) = \frac{\Delta_k}{\Delta_{\text{ideal}}} - 1.$$

INL represents the cumulative deviation of the actual transfer curve from the best-fit straight line across the full code range.[FN1]

In the context of IM-DD PAM and BPAM, non-zero DNL and INL distort the intended amplitude levels: some constellation points are compressed, others expanded or shifted, and the mapping from digital word (the N-bit code fed to the DAC at each sampling instant) to optical power becomes pattern-dependent. This manifests as:

- Increased EVM, even in the absence of additive noise.
- Level-dependent eye closure (some eyes degrade more than others).
- Generation of deterministic distortion spurs in the electrical spectrum.

Analytical and numerical studies of PAM4 transmitters show that realistic INL/DNL profiles can contribute several tenths of a dB to the implementation penalty at a given BER, especially when combined with bandwidth limitations and

pre-emphasis¹⁶. For BPAM, where the sign of the waveform carries additional information, asymmetric nonlinearities between positive and negative codes are particularly harmful and must be accurately captured in a surrogate model.

3.2.1.3 Dynamic errors and glitches

Static linearity metrics such as INL and DNL do not provide a complete picture. At high symbol rates, *dynamic errors* become dominant. In segmented current-steering DACs, major-bit transitions (e.g., from code 0111... to 1000...) rarely switch all devices instantaneously. Mismatched switching times and finite output settling give rise to *glitches*: short voltage or current spikes superimposed on the desired waveform¹⁷. Glitch energy is powerful when many bits toggle simultaneously and when the output drives a wideband load, such as a driver input or a transmission line.

In the time domain, glitches appear as overshoot, undershoot, or ringing at transitions, effectively adding high-frequency noise and intersymbol interference (ISI). In the frequency domain, they populate the spectrum with additional high-frequency components that can be partially filtered by the analog front end but still degrade eye diagrams. For multilevel PAM and BPAM, where precise transition timing and level separation are critical, such dynamic non-idealities can significantly increase EVM and contribute to error floors. A realistic digital surrogate must therefore model not only the static transfer curve but also the finite-time response of the DAC to code changes.

3.2.1.4 Thermal and reference noise

Beyond quantization, DAC outputs are contaminated by *thermal noise* from resistive elements, switches, and output buffers, as well as by noise in reference voltages or currents. These contributions are often modeled as additive Gaussian noise at the DAC output, with a power spectral density set by the equivalent output

¹⁶ M. Nazarathy and I. Tomkos, “‘Perfect’ PAM4 serial digital-optical conversion,” IEEE Photonics Technology Letters, vol. 33, no. 9, pp. 475-478, 2021.

¹⁷ X. Li & L. Zhou, “A survey of high-speed and high-resolution current-steering DACs,” Journal of Semiconductors, vol. 41, no. 11, art. 111404, 2020.

resistance and the noise bandwidth of the reconstruction path¹⁸. In many IM-DD transmitter designs, especially when operating near full scale, quantization noise dominates. However, as resolution increases (e.g., with 8-10-bit DACs and strong digital preprocessing), thermal and reference noise can become comparable to, or exceed, the ideal quantization floor.

From a system viewpoint, thermal noise at the DAC output adds directly to other electrical noise sources (driver and amplifier noise). It propagates through the modulator, appearing as additional intensity noise at the photodiode. In BPAM systems, where sign reconstruction relies on subtle temporal correlations in the waveform, this additional electrical noise can erode the margin relative to the theoretical BPAM performance.

3.2.1.5 Timing jitter

Finally, *clock jitter* introduces timing uncertainty in the DAC's sampling instants. If the ideal output is $x(t)$ and the actual DAC updates occur at $t_k + \delta t_k$, small random perturbations δt_k translate into amplitude errors approximately equal to $\dot{x}(t_k)\delta t_k$. For a tone of frequency f , this yields an approximate jitter-induced SNR:

$$\text{SNR}_{\text{jitter}} \approx -20 \log_{10} (2\pi f \sigma_t),$$

where σ_t is the root-mean-square timing jitter¹⁹. This expression, originally derived for high-speed data-converter and sampling systems, highlights that jitter penalties increase with both signal frequency content and jitter variance.

In IM-DD PAM and BPAM waveforms, the relevant “frequency” is linked to edge rates and pre-emphasized transitions rather than to a single carrier. Strong digital pre-emphasis and high baud rates increase the slew rate $|\dot{x}(t)|$, making the link more sensitive to jitter. At the optical level, jitter shows up as horizontal eye

¹⁸ L. Michaeli and J. Šaliga, “Error models of the analog to digital converters,” *Measurement Science Review*, vol. 14, no. 2, pp. 62-77, 2014.

¹⁹ N. Da Dalt, M. Harteneck, C. Sandner, and A. Wiesbauer, “On the jitter requirements of the sampling clock for analog-to-digital converters,” *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 49, no. 9, pp. 1354-1360, Sept. 2002.

closure and increased EVM. Significantly, jitter interacts with INL/DNL and bandwidth limitations: distorted waveforms often have steeper or more irregular transitions, which magnify the effective amplitude error for a given σ_t ²⁰.

3.2.1.6 Implications for the digital surrogate

Most link-level IM-DD studies model the DAC as an ideal quantizer with additive Gaussian noise, or even as a continuous-time linear block. More detailed converter literature, on the other hand, provides rich error taxonomies and characterization methods but seldom connects them to system-level PAM/BPAM performance. A central objective of this thesis is to bridge this gap: the digital surrogate explicitly represents quantization, INL/DNL, glitch-like dynamic behavior, thermal noise, and jitter as separate, parametrized impairments at the DAC stage. This allows us to:

- Study how each impairment class contributes to BER, EVM, and eye-diagram metrics for PAM and BPAM.
- Calibrate DAC models against measurements from real hardware.
- Provide ML-ready metadata linking device-level parameters (e.g., INL profile, jitter RMS) to system-level performance.

In this way, the DAC block in the surrogate becomes not just a “bit-to-voltage” converter, but a realistic, tunable model of a key bottleneck in high-speed IM-DD transmitters.

3.2.2 Semiconductor optical amplifier (SOA): gain saturation, pattern effects, and ASE noise

Semiconductor optical amplifiers (SOAs) are widely used in optical communication systems as compact, electrically pumped gain elements that can be deployed either as discrete components or integrated on photonic platforms. In IM-DD links, they are especially attractive when *compactness*, *integrability*, and *wavelength flexibility* are priorities. However, their use introduces characteristic *noise* and *nonlinear distortions* that become increasingly relevant at high symbol

²⁰ M. Nazarathy and I. Tomkos, “‘Perfect’ PAM4 serial digital-optical conversion,” IEEE Photonics Technology Letters, vol. 33, no. 9, pp. 475-478, 2021

rates and with multilevel modulation (PAM4/6/8) or emerging bipolar formats (BPAM)²¹. From the viewpoint of a digital surrogate, the SOA cannot be treated as a generic “lumped gain” block: realistic emulation requires capturing (i) gain saturation and gain dynamics, (ii) ASE noise generation, and (iii) amplitude-phase coupling and fast intraband nonlinearities, because these impairments directly shape BER/EVM and can generate strong data-dependence that is critical for ML training²².

3.2.2.1 SOA roles in IM/DD systems: booster vs pre-amplifier (and “passive” use)

In practice, SOAs appear in IM/DD systems in two primary amplifier roles:

(a) Booster amplifier (Tx-side). A booster SOA is placed after the transmitter (e.g., after an EML/MZM or an integrated modulator) to increase the launched optical power and extend the link power budget. This is particularly relevant when the transmitter output power is limited by integration constraints, thermal limits, or eye-safety regulations²³.

(b) Pre-amplifier (Rx-side). A pre-amplifier SOA is placed before the photodiode/TIA to improve receiver sensitivity by increasing the detected signal current. This usage is common in reach-extended IM/DD, and in some passive optical network PON and mobile fronthaul architectures, where improved sensitivity directly translates into higher split ratios or longer reach²⁴.

In addition, SOA sections can be used in a “passive” sense (biased off or below transparency) as lossy/absorptive elements in gates/switches or as part of integrated

²¹ A. Sobhanan, A. Anthur, S. O’Duill, M. Pelusi, S. Namiki, L. Barry, D. Venkitesh, and G. P. Agrawal, “Semiconductor optical amplifiers: recent advances and applications”, *Advances in Optics and Photonics*, vol. 14, no. 3, pp. 571-651, 2022.

²² *Ibidem*.

²³ Z. Li, L. Yi, and W. Hu, “Symmetric 40-Gb/s TWDM-PON with 51-dB loss budget by using a single SOA as preamplifier, booster and format converter in ONU”, *Optics Express*, vol. 22, no. 20, pp. 24398-24404, 2014.

²⁴ P. Zhu, Y. Yoshida, K. Akahane, and K. Kitayama, “High-speed reach-extended IM-DD system with low-complexity DSP for 6G fronthaul [Invited]”, *Journal of Optical Communications and Networking*, vol. 16, pp. A24-A32, 2023.

photonic subsystems. While this is not the primary focus here, it matters to a surrogate because bias conditions determine whether the block behaves as an amplifier, a nonlinear absorber, or an intermediate regime²⁵.

3.2.2.2 Small-signal gain and gain saturation (static compression)

A first-order SOA description begins with the *small-signal (unsaturated) gain* G_0 and a *saturation parameter* (often expressed via saturation power P_{sat} or saturation energy E_{sat}). Under steady-state assumptions, gain compression is frequently represented by a simplified relationship such as:

$$G(P_{\text{in}}) \approx \frac{G_0}{1 + P_{\text{in}}/P_{\text{sat}}}.$$

Although this static form is insufficient at high speeds, it explicitly highlights an essential point for IM/DD multilevel signaling: higher optical peak power (or higher PAPR) pushes the SOA into compression, reducing the effective distance between PAM levels, increasing EVM, and creating format-dependent implementation penalties²⁶.

For the surrogate, the minimum “static” parameter set should therefore include G_0 , $P_{\text{sat}}/E_{\text{sat}}$, and a calibrated operating point (bias current, temperature), because these determine the degree of amplitude compression experienced by PAM/BPAM waveforms²⁷.

3.2.2.3 Gain dynamics and pattern effects (memory and data dependence)

At high symbol rates, SOA behavior is dominated not only by static compression but also by gain recovery dynamics. Carrier depletion and recovery introduce a memory effect that makes the instantaneous gain depend on recent

²⁵ A. Sobhanan, A. Anthur, S. O’Duill, M. Pelusi, S. Namiki, L. Barry, D. Venkitesh, and G. P. Agrawal, “Semiconductor optical amplifiers: recent advances and applications”, *Advances in Optics and Photonics*, vol. 14, no. 3, pp. 571-651, 2022.

²⁶ M. S. Wartak, “*Computational Photonics: An Introduction with MATLAB*”, Cambridge University Press, Cambridge, Ch. 8 (gain saturation model), 2013.

²⁷ R. Dionisio, G. Parca, C. Reis, & A. Teixeira. “Operational parameter optimisation of MZI-SOA using multi-objective genetic algorithms”. *Electronics Letters*, 47, 561-562. 2011

symbol history, producing *pattern effects* (sometimes described as “patterning”): identical symbols can be amplified differently depending on preceding levels²⁸.

A commonly used dynamic abstraction expresses gain evolution with a recovery time constant τ_c and a saturation term driven by the input optical power $P_{\text{in}}(t)$, e.g.:

$$\frac{dG(t)}{dt} = \frac{G_0 - G(t)}{\tau_c} - \frac{G(t)P_{\text{in}}(t)}{E_{\text{sat}}}.$$

In multilevel PAM, this translates into *level-dependent distortion*, and *eye closure* that cannot be fully corrected by linear equalization alone, particularly when τ_c is comparable to (or longer than) the symbol period²⁹. For BPAM-like schemes, where the receiver may rely on oversampling and reconstruction mechanisms, SOA-induced memory can also perturb the waveform structure that reconstruction depends on, making accurate SOA modeling valuable even when the optical front-end remains “IM/DD-simple.”

From a surrogate/ML standpoint, pattern effects are not a nuisance to be averaged out: they are a *structured source of nonlinearity* that can be exploited (or must be mitigated) via data-driven compensation. This motivates logging internal variables (or proxies) such as *instantaneous gain state*, *effective compression factor*, and *recovery-time parameters* as part of the metadata produced for training³⁰.

3.2.2.4 ASE noise and OSNR penalties in direct detection

SOAs generate *amplified spontaneous emission (ASE)*, which degrades SNR/OSNR and, under direct detection, produces characteristic beating terms after

²⁸ A. Sobhanan, A. Anthur, S. O’Duill, M. Pelusi, S. Namiki, L. Barry, D. Venkitesh, and G. P. Agrawal, “Semiconductor optical amplifiers: recent advances and applications”, *Advances in Optics and Photonics*, vol. 14, no. 3, pp. 571-651, 2022.

²⁹ F. Ginovart, J.C. Simon, I. Valiente, “Gain recovery dynamics in semiconductor optical amplifier”. *Optics Communications*, 199, 111-115. 2001.

³⁰ A. Omid, M. Banawan, E. Weckenmann, B. Paquin, A. Geravand, Z. Zheng, “Auto-Encoder Optimized PAM IM/DD Transceivers for Amplified Fiber Links,” *Journal of Lightwave Technology*, vol. 43, pp. 547-556, 2024.

the photodiode. A widely used expression for the *output ASE power* in an optical bandwidth B_0 is:

$$P_{\text{ASE}} \approx 2, n_{\text{sp}}, h\nu, (G - 1)B_0,$$

where n_{sp} is the spontaneous emission factor, $(h\nu)$ is the photon energy, and the factor 2 accounts for polarization modes³¹. The SOA *noise figure* is closely tied to n_{sp} (and the operating point), and in high-gain regimes, it is often expressed in terms of n_{sp} to link device physics to link-level OSNR budgeting³².

After square-law detection, ASE contributes not only as added optical noise but through *signal-ASE beat noise* and *ASE-ASE beat noise*, which can be significant penalties in pre-amplified IM/DD receivers. These terms scale with both signal power and ASE power, which is why the SOA operating point (gain/compression) is inseparable from noise performance in system design³³.

Experimentally, SOA pre-amplification has been demonstrated to enable high-rate PAM4 IM/DD in access-style architectures; for example, SOA pre-amplified 100 Gb/s/ λ PAM4 downstream transmission in an O-band TDM-PON context highlights both the sensitivity benefit and the practical need to manage ASE and nonlinear penalties³⁴.

3.2.2.5 Nonlinearities most relevant to IM/DD PAM and BPAM

Beyond gain saturation and slow gain recovery, several SOA nonlinear mechanisms can matter in high-speed IM/DD, especially when bandwidth is tight, PAM order increases, or the receiver relies on sophisticated DSP (as in BPAM-compatible processing):

³¹ A. Sobhanan, A. Anthur, S. O'Duill, M. Pelusi, S. Namiki, L. Barry, D. Venkitesh, and G. P. Agrawal, "Semiconductor optical amplifiers: recent advances and applications", *Advances in Optics and Photonics*, vol. 14, no. 3, pp. 571-651, 2022.

³² Giuliani, G., & D'Alessandro, D. "Noise analysis of conventional and gain-clamped semiconductor optical amplifiers." *Journal of Lightwave Technology*, 18, 1256-1263, 2000.

³³ Derevyanko, S., & Turitsyn, S. "Bit-Error Probability for Direct Detection of Optical RZ Signal Degraded by ASE Noise and Timing Jitter." *Journal of Lightwave Technology*, 25, 638-643, 2007.

³⁴ J. Zhang, J. Yu, J. S. Wey, X. Li, L. Zhao, K. Wang, M. Kong, W. Zhou, J. Xiao, X. Xin, & F. Zhao, "SOA Pre-Amplified 100 Gb/s/ λ PAM-4 TDM-PON Downstream Transmission Using 10 Gbps O-Band Transmitters", *Journal of Lightwave Technology*, vol. 38, pp. 185-193, 2020.

- **Carrier heating (CH) and spectral hole burning (SHB):** ultrafast intraband effects that reshape gain on sub-ps to ps scales, contributing to nonlinear distortion at very high symbol rates³⁵.
- **Amplitude-phase coupling (linewidth enhancement factor α):** power-dependent refractive index changes produce phase modulation (“chirp”) accompanying amplitude changes. Even in IM/DD, this can lead to amplitude distortion due to filtering and dispersion (and it matters whenever the effective field waveform structure is essential for DSP reconstruction)³⁶.
- **Four-wave mixing (FWM) and cross-gain modulation (XGM):** particularly relevant in WDM or multi-carrier contexts, but also as secondary distortions in dense spectral packing³⁷.
- **Polarization dependence (PDG/PMD sensitivity) and bias-dependent dynamics:** practical non-idealities that can introduce slow drifts or operating-point dependence, relevant for robust surrogate calibration³⁸.

For a digital surrogate intended for end-to-end optimization and ML training, the implication is clear: the SOA block should expose a tunable model hierarchy, ranging from (i) gain + ASE only, up to (ii) gain dynamics + ASE beating, and further to (iii) amplitude-phase coupling and fast intraband terms, so that fidelity can be increased when the optimization task demands it³⁹.

3.2.2.6 O-band vs C-band use, and a short comparison with EDFA

O-band vs C-band. In many short-reach and access-style IM/DD scenarios, O-band operation (around 1310 nm) is attractive due to low chromatic dispersion in standard fiber and the ecosystem of cost-effective components, while C-band

³⁵ A. Marculescu *et al.*, “Spectral signature of nonlinear effects in semiconductor optical amplifiers”, *Optics Express*, vol. 25, no. 24, pp. 29526-29559, 2017.

³⁶ M. J. Connelly, “*Semiconductor Optical Amplifiers*”, Springer, 2002.

³⁷ K. E. Stubkjaer, “Semiconductor optical amplifier-based all-optical gates for high-speed optical processing”, *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 6, no. 6, 2000.

³⁸ M. J. Connelly, “*Semiconductor Optical Amplifiers*”, Springer, 2002.

³⁹ A. Omid, M. Banawan, E. Weckenmann, B. Paquin, A. Geravand, Z. Zheng, “Auto-Encoder Optimized PAM IM/DD Transceivers for Amplified Fiber Links,” *Journal of Lightwave Technology*, vol. 43, pp. 547-556, 2024.

operation (around 1550 nm) is dominant in long-haul/metro due to mature amplification infrastructure and WDM component availability⁴⁰. SOAs are applicable across both wavelength regions, and their appeal is strongest when *integration and compact amplification* are desired; conversely, their penalties (ASE plus nonlinear dynamics) become more pronounced as baud rate and PAM order increase. Recent IM/DD literature on reach extension and fronthaul highlights the continuing relevance of optical amplification (including SOA-like preamplification concepts) alongside low-complexity DSP approaches⁴¹.

SOA vs EDFA (high level). Compared with EDFAs, SOAs are generally more compact and easier to integrate. They can be engineered across broader wavelength regions, but they typically exhibit stronger nonlinear dynamics and different noise/linearity trade-offs that are especially important under direct detection⁴². For the surrogate, this means that an “amplifier” cannot be treated generically: SOA-specific dynamics and distortion mechanisms must be represented explicitly whenever the experimental system uses an SOA (either as booster or pre-amplifier), because these effects shape the statistics of the waveform that downstream DSP/ML will see⁴³.

3.2.3 Photodiode / TIA and ADC (bandwidth, noise, sampling, quantization)

In IM-DD links, the receiver front-end typically consists of an *optical-to-electrical conversion chain* (photodiode plus transimpedance amplifier), followed by *sampling and digitization* (ADC), and then receiver DSP. Even when the optical

⁴⁰ X. Zhou, R. Urata, and H. Liu, “Beyond 1 Tb/s Intra-Data Center Interconnect Technology: IM-DD OR Coherent?”, *Journal of Lightwave Technology*, vol. 38, pp. 475-484, 2020.

⁴¹ T. Wettlin, S. Calabrò, T. Rahman, J. Wei, N. Stojanovic, S. Pachnicke, “DSP for High-Speed Short-Reach IM/DD Systems Using PAM”, *Journal of Lightwave Technology*, vol. 38, pp. 6771-6778, 2020.

⁴² A. Sobhanan, A. Anthur, S. O’Duill, M. Pelusi, S. Namiki, L. Barry, D. Venkitesh, and G. P. Agrawal, “Semiconductor optical amplifiers: recent advances and applications”, *Advances in Optics and Photonics*, vol. 14, no. 3, pp. 571-651, 2022.

⁴³ B. Omid, E. P. da Silva, B. Neto, L. G. D. A. Lemos, I. M. Moroz, M. P. Yankov, E. Agrell, P. Bayvel, and R. I. Killey, “Auto-Encoder Optimized PAM IM/DD Transceivers for Amplified Fiber Links”, *Journal of Lightwave Technology*, vol. 43, pp. 547-556, 2024.

path is “simple” (direct detection), the *electrical bandwidth, noise, and timing/quantization limits* of the PD/TIA/ADC chain strongly shape the achievable BER/EVM for multilevel PAM. They become even more critical when the receiver relies on oversampling and waveform structure (e.g., BPAM-style reconstruction). This subsection summarizes the dominant PD/TIA and ADC impairments and highlights which parameters must be explicitly represented in a digital surrogate intended for realistic end-to-end evaluation.

3.2.3.1 Photodetection and electrical front-end: roles and interfaces

A photodiode (PD) converts received optical power $P(t)$ into photocurrent $i(t)$ approximately as:

$$i(t) = \mathcal{R}P(t) + i_{\text{dark}} + n_{\text{shot}}(t),$$

where $\mathcal{R}$ is the responsivity (A/W), i_{dark} is the dark current, and $n_{\text{shot}}(t)$ models shot-noise fluctuations⁴⁴. In short-reach IM-DD receivers, *PIN photodiodes* are common due to their linearity and bandwidth; avalanche photodiodes (APDs) trade higher sensitivity for excess noise and complexity⁴⁵.

The transimpedance amplifier (TIA) converts current to voltage, providing gain and defining a large fraction of the *receiver’s electrical transfer function* and noise floor. At the system level, PD + TIA (and any following analog equalization) are well represented as a band-limited linear front-end with frequency response $H_{\text{rx}}(f)$, plus additive noise terms referenced to either input current or output voltage. This $H_{\text{rx}}(f)$ directly shapes eye opening, ISI, and the distortion seen by the ADC/DSP⁴⁶.

⁴⁴ G. P. Agrawal, “*Fiber-Optic Communication Systems*,” Wiley, 5th ed., 2021.

⁴⁵ *Ibidem*.

⁴⁶ K. Zhong, X. Zhou, and J. Huo, “*Digital signal processing for short-reach optical communications: A review of current technologies and future trends*,” Journal of Lightwave Technology, vol. 36, pp. 377-400, 2018.

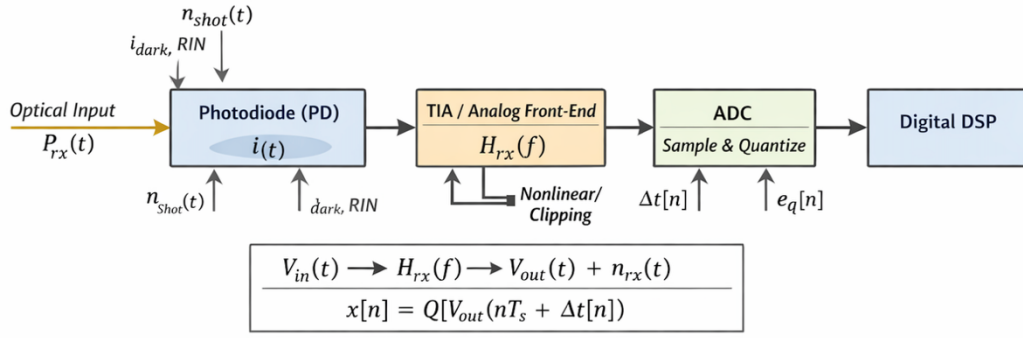


Figure 3.2 Illustration of the receiver front-end model with equivalent transfer function and noise

3.2.3.2 Noise mechanisms and sensitivity limits

For IM-DD, receiver performance is often limited by a combination of *shot noise*, *thermal/electronic noise*, and (when present) *intensity noise* sources (laser relative intensity noise or RIN, amplifier-related beating terms in amplified links, etc.). A widely used base model expresses the total detected-noise variance in an equivalent electrical bandwidth B as the sum of contributions⁴⁷, e.g.:

- **Shot noise (signal-dependent):** $\sigma_{shot}^2 \approx 2qI_{avg}B$, where q is the electron charge and I_{avg} includes signal and background currents.
- **Thermal/TIA noise (often signal-independent in first order):** expressed via input-referred current noise density or an equivalent thermal-noise term integrated over bandwidth⁴⁸.

For multilevel PAM, these noise sources do not merely “raise BER”; they *compress effective level separations* (eye heights between adjacent PAM levels), so that the same incremental noise penalty translates into *disproportionately larger BER/EVM penalties* at higher modulation orders⁴⁹. In practice, receiver sensitivity

⁴⁷ X. Huang, et al. “Optimized equalization based on MAP detection for C-band bandwidth-limited IM/DD systems.” *Optics letters*, 49 13, 3644-3647, 2024 .

⁴⁸ G. P. Agrawal, *"Fiber-Optic Communication Systems"*, Wiley, 5th ed, 2021.

⁴⁹ K. Zhong, X. Zhou, and J. Huo, *"Digital signal processing for short-reach optical communications: A review of current technologies and future trends"*, *Journal of Lightwave Technology*, vol. 36, pp. 377-400, 2018.

and the feasible link budget depend strongly on the combined PD/TIA noise and bandwidth profile, not just on optical power alone⁵⁰.

3.2.3.3 Bandwidth limitations, ISI, and front-end linearity

At high baud rates, limited PD/TIA bandwidth causes *pulse spreading and intersymbol interference (ISI)*. For PAM4/PAM8 (and BPAM variants), bandwidth limitation reduces vertical eye openings. It introduces pattern-dependent distortions that increase EVM and degrade BER, often requiring receiver equalization (linear FFE/DFE, MLSE, or other DSP) to recover performance⁵¹. Practical short-reach IM/DD studies repeatedly emphasize that system design is *co-limited* by (i) available transmitter bandwidth (DAC/driver/modulator), and (ii) receiver bandwidth and noise (PD/TIA/ADC), with the DSP positioned to compensate residual ISI within feasible complexity⁵².

Front-end *linearity and clipping* also matter: the TIA/AFE must remain sufficiently linear over the amplitude distribution of the detected waveform. With multilevel formats, any soft-compression or hard-clipping creates level-dependent errors (outer levels distort first), which increase EVM and yield BER floors that cannot be “averaged out” by increasing the number of samples⁵³.

3.2.3.4 ADC sampling rate, quantization, and ENOB constraints

After the analog front end (AFE), the ADC samples the waveform at F_s . In short-reach receivers, F_s is often chosen to support timing recovery and equalization. For BPAM-like approaches, *intentional oversampling* can be part of

⁵⁰ X. Zhou, R. Urata, and H. Liu, "Beyond 1 Tb/s Intra-Data Center Interconnect Technology: IM-DD OR Coherent?", Journal of Lightwave Technology, vol. 38, pp. 475-484, 2020.

⁵¹ K. Zhong, X. Zhou, and J. Huo, "Digital signal processing for short-reach optical communications: A review of current technologies and future trends", Journal of Lightwave Technology, vol. 36, pp. 377-400, 2018.

⁵² T. Wettlin, , S. Calabrò, T. Rahman, J. Wei, N. Stojanovic, S. Pachnicke, "DSP for High-Speed Short-Reach IM/DD Systems Using PAM", Journal of Lightwave Technology, vol. 38, pp. 6771-6778, 2020.

⁵³ *Ibidem* 51.(K. Zhong et al.)

the information-recovery mechanism, so the ADC becomes structurally coupled to the modulation concept, not just a measurement device⁵⁴.

Quantization imposes a finite amplitude resolution. For a uniform (N)-bit quantizer with full-scale range V_{FS} , the step is $\Delta = V_{FS}/2^N$. Under the common additive quantization noise approximation, for sufficiently "busy" signals, the variance of the quantization noise is given by $\sigma_q^2 \approx \Delta^2/12$, and the ideal quantization-limited SNR scales approximately as $6.02N + 1.76$ dB (for a full-scale sinusoid)⁵⁵. While real PAM waveforms are not full-scale sinusoids, this relation remains a practical first-order guide: higher-order PAM and BPAM-like signaling typically require tighter control of ADC resolution, scaling, and clipping to preserve the effective distance between decision levels and to avoid constellation-dependent distortion⁵⁶.

A further practical limitation is that high-speed ADCs have an effective number of bits (ENOB) that typically degrades as sampling rate and input bandwidth increase, making "more samples" and "more bits" competing requirements in real implementations, precisely the kind of hardware trade-off a surrogate should be able to emulate when evaluating system-level DSP/ML strategies⁵⁷.

3.2.3.5 Sampling-time errors: clock jitter and aperture uncertainty

Timing imperfections in sampling (clock jitter, aperture uncertainty, interleaving skew) produce amplitude errors that scale with the *slope* of the received waveform. For high-frequency components, even slight RMS time jitter σ_t can impose a strong SNR limit; a commonly used approximation expresses jitter-limited SNR as

⁵⁴ *Ibidem*.

⁵⁵ W. Kester (Ed.), *"The Data Conversion Handbook"*, Newnes/Elsevier, ISBN: 978-0-7506-7841-4, 2005.

⁵⁶ T. Wettlin, , S. Calabrò, T. Rahman, J. Wei, N. Stojanovic, S. Pachnicke, *"DSP for High-Speed Short-Reach IM/DD Systems Using PAM"*, Journal of Lightwave Technology, vol. 38, pp. 6771-6778, 2020.

⁵⁷ X. Zhou, R. Urata, and H. Liu, *"Beyond 1 Tb/s Intra-Data Center Interconnect Technology: IM-DD OR Coherent?"*, Journal of Lightwave Technology, vol. 38, pp. 475-484, 2020.

$$\text{SNR}_{\text{jitter}} \approx -20 \log_{10} (2\pi f_{\text{in}} \sigma_t),$$

highlighting that jitter becomes increasingly harmful as symbol rates and waveform bandwidth grow⁵⁸. In multilevel IM-DD links, this manifests as increased EVM and timing-dependent decision errors; for BPAM-like schemes that rely on oversampled reconstruction, sampling-time errors can also disrupt the consistency of the reconstruction operation (because it depends on the relative timing of samples within the symbol interval)⁵⁹.

3.2.3.6 Implications for a digital surrogate and ML-ready metadata

For a surrogate that is meant to both (i) faithfully reproduce measured end-to-end behavior and (ii) generate training data for ML-based optimization, PD/TIA/ADC modeling should expose both physical parameters and measurable aggregates. A minimal, surrogate-friendly parameterization includes:

- **PD:** responsivity $\mathcal{R}$, bandwidth (or impulse response), dark current, optional RIN/beat-noise hooks for amplified cases⁶⁰.
- **TIA/AFE:** transimpedance gain, effective $H_{\text{rx}}(f)$, input-referred noise density (or equivalent integrated noise), linear range/clipping model⁶¹.
- **ADC:** (F_s), effective resolution/ENOB, quantizer type and scaling, sampling jitter (σ_t), and optional interleaving mismatch terms if required by the targeted hardware class⁶².

Crucially, the surrogate should also output ML-useful metadata such as per-block SNR/EVM contributions, clipping statistics, and bandwidth/ISI indicators, enabling learning algorithms to distinguish “noise-limited” vs “bandwidth-limited”

⁵⁸ W. Kester (Ed.), *"The Data Conversion Handbook"*, Newnes/Elsevier, ISBN: 978-0-7506-7841-4, 2005.

⁵⁹ J. G. Proakis, M. Salehi, *"Digital Communications"*, McGraw-Hill, 5th ed., ISBN: 978-0-07-295716-7, 2008.

⁶⁰ B. E. A. Saleh and M. C. Teich, *"Fundamentals of Photonics"*, Wiley, 3rd ed., 2019.

⁶¹ K. Zhong, X. Zhou, and J. Huo, *"Digital signal processing for short-reach optical communications: A review of current technologies and future trends"*, Journal of Lightwave Technology, vol. 36, pp. 377-400, 2018.

⁶² W. Kester (Ed.), *"The Data Conversion Handbook"*, Newnes/Elsevier, ISBN: 978-0-7506-7841-4, 2005.

regimes and to optimize transmitter DSP and receiver DSP under realistic front-end constraints⁶³.

3.3 Existing modeling and simulation approaches

3.3.1 Commercial and academic link simulators, analytical models, vendor tools

State-of-the-art IM-DD link design is rarely performed using a single “monolithic” model. Instead, the literature converges on a *multi-layer modeling stack* that trades fidelity against runtime depending on the question being asked (architecture screening, DSP feasibility, hardware budgeting, or final performance validation). In practice, the dominant approaches can be grouped into (i) analytical and semi-analytical link models, (ii) system-level time-domain simulators (commercial or academic), and (iii) device-/circuit-aware co-simulation and “vendor-tool” workflows that import measured component responses (e.g., S-parameters, transfer functions, noise figures) into end-to-end evaluations^{64,65}.

3.3.1.1 Analytical and semi-analytical link models (fast screening)

The first layer consists of *closed-form models*, in which key link quantities are given by explicit analytical expressions, and *semi-closed-form models*, in which the main relationships remain analytical but include a small number of numerically evaluated or empirically fitted terms; together, they provide rapid engineering insight and enable efficient parameter sweeps. In IM-DD, this typically includes: (a) link-budget style calculations (loss, gain, OSNR/SNR mappings), (b) simplified receiver noise models (shot/thermal/RIN and ASE when amplification is present), and (c) approximate BER/EVM relations under assumptions such as Gaussian noise

⁶³ T. Wettlin, , S. Calabrò, T. Rahman, J. Wei, N. Stojanovic, S. Pachnicke, “*DSP for High-Speed Short-Reach IM/DD Systems Using PAM*”, Journal of Lightwave Technology, vol. 38, pp. 6771-6778, 2020.

⁶⁴ M. S. Wartak, “*Computational Photonics: An Introduction with MATLAB*,” Cambridge University Press, Cambridge, 2013.

⁶⁵ T. Wettlin; S. Calabrò; T. Rahman; J. Wei; N. Stojanovic, “*DSP for High-Speed Short-Reach IM/DD Systems Using PAM*”, Journal of Lightwave Technology, vol. 38, pp. 6771-6778, 2020.

at the decision variable^{66,67}. These models are especially valuable in early-stage trade studies (e.g., “IM-DD vs coherent” feasibility at a given lane rate, reach, and cost envelope) because they expose which constraints dominate (bandwidth, noise, linearity, or DSP complexity) without committing to a full waveform simulation⁶⁸.

Even when end-to-end simulation is eventually performed, analytical results remain important as *sanity checks* and as anchors for parameter calibration. For example, when SOA pre-amplification is considered, ASE-driven sensitivity limits and timing-jitter penalties can be estimated analytically and compared with time-domain Monte Carlo results to confirm that the simulator’s noise-injection and filtering assumptions are consistent⁶⁹.

3.3.1.2 System-level time-domain simulators (waveform-accurate link evaluation)

The second layer is *system-level waveform simulation*, in which the transmitter and receiver are modeled as a chain of blocks that operate on sampled signals (electrical waveform, optical power/field proxy, and detected current/voltage). This level is the workhorse for IM-DD PAM research, because it naturally captures: (i) bandwidth limitations and ISI, (ii) nonlinear transfer functions and clipping, (iii) quantization and mixed-signal artifacts (DAC/ADC), and (iv) the interaction between DSP and hardware impairments⁷⁰.

A typical short-reach IM-DD PAM workflow is: define the symbol mapping and pulse shaping; set an oversampling factor; apply bandwidth-limited analog

⁶⁶ S. Derevyanko and S. Turitsyn, “Bit-Error Probability for Direct Detection of Optical RZ Signal Degraded by ASE Noise and Timing Jitter,” *Journal of Lightwave Technology*, vol. 25, pp. 638-643, 2007.

⁶⁷ X. Zhou, R. Urata, and H. Liu, “Beyond 1 Tb/s Intra-Data Center Interconnect Technology: IM-DD OR Coherent?,” *Journal of Lightwave Technology*, vol. 38, pp. 475-484, 2020.

⁶⁸ *Ibidem*.

⁶⁹ S. Derevyanko and S. Turitsyn, “Bit-Error Probability for Direct Detection of Optical RZ Signal Degraded by ASE Noise and Timing Jitter,” *Journal of Lightwave Technology*, vol. 25, pp. 638-643, 2007.

⁷⁰ T. Wettlin; S. Calabrò; T. Rahman; J. Wei; N. Stojanovic, “DSP for High-Speed Short-Reach IM/DD Systems Using PAM”, *Journal of Lightwave Technology*, vol. 38, pp. 6771-6778, 2020.

front-end responses; inject noise sources at physically meaningful points; quantize where appropriate; and finally run receiver DSP (timing recovery, equalization, slicing) and estimate metrics such as BER/EVM/eye opening through Monte Carlo over many blocks⁷¹. Importantly, this approach is **format-agnostic**: the same simulator can be used for conventional unipolar PAM as well as emerging variants (e.g., shaped PAM), as long as the correct front-end and DSP assumptions are implemented⁷².

In reach-extended scenarios (e.g., fronthaul, PON-like budgets, amplified IM-DD), system-level simulation becomes even more central because it must jointly represent optical amplification, receiver sensitivity, and DSP feasibility under tight complexity constraints. Studies on practical reach extension for IM-DD explicitly rely on such block-based simulation to quantify the performance/complexity trade-offs of different DSP and transceiver options⁷³. Likewise, experimental IM-DD demonstrations that employ SOA pre-amplification typically pair measurements with system simulations (or semi-analytical baselines) to separate device physics limitations from implementation penalties (filtering, quantization, equalizer mismatch)⁷⁴.

3.3.1.3 Commercial simulators and vendor-tool workflows (component libraries and measurement-driven modeling)

Commercial optical-link simulators are widely used in both industry and academia primarily because they provide (a) *validated component libraries* (lasers, modulators, SOAs/EDFAs, fiber/channel blocks, photodiodes, electrical filters), (b) *visual system composition* and parameter sweeps, and (c) optional interfaces for co-simulation with external DSP environments. In practice, these tools are strongest

⁷¹ *Ibidem*.

⁷² P. Zhu, Y. Yoshida, K. Akahane, and K. Kitayama, “High-speed reach-extended IM-DD system with low-complexity DSP for 6G fronthaul [Invited],” *Journal of Optical Communications and Networking*, vol. 16, pp. A24-A32, 2023.

⁷³ *Ibidem*.

⁷⁴ J. Zhang, *et al.* “SOA Pre-Amplified 100 Gb/s/λ PAM-4 TDM-PON Downstream Transmission Using 10 Gbps O-Band Transmitters”, *Journal of Lightwave Technology*, vol. 38, pp. 185-193, 2020.

when the design question is “*system integration*” (budgets, margins, sensitivity to component tolerances) rather than the invention of novel DSP or the modeling of a very specific, lab-calibrated impairment.

Vendor-tool workflows also commonly incorporate *measured component characterizations*, e.g., importing frequency responses to obtain an equivalent transfer function, using specified gain/NF values for amplifiers, or fitting a compact model to experimental curves. This “measurement-driven modeling” is particularly relevant for SOAs, where gain compression, ASE, and recovery dynamics depend strongly on bias and operating point; modern SOA reviews summarize both the physical mechanisms and the compact models that system simulators typically implement⁷⁵. Noise modeling in such tools often traces back to well-established analytical treatments of SOA ASE and noise statistics (including gain-clamped vs conventional operation), which then become parameterized blocks inside a larger link simulation⁷⁶.

3.3.1.4 Academic simulators and research codebases (flexibility and reproducibility)

Alongside commercial platforms, a large fraction of the state-of-the-art literature relies on *custom MATLAB/Python simulators*, especially when the contribution requires: (i) non-standard modulation formats or reconstruction assumptions, (ii) detailed mixed-signal modeling (DAC/ADC artifacts), (iii) custom equalization/ML blocks, or (iv) extraction of rich metadata beyond standard BER/EVM (e.g., per-block distortion fingerprints needed for learning-based optimization). A computational-photonics-style approach, which explicitly implements propagation/filtering models and device blocks within a scripting

⁷⁵ A. Sobhanan, A. Anthur, S. O’Duill, M. Pelusi, S. Namiki, L. Barry, D. Venkitesh, and G. P. Agrawal, “*Semiconductor optical amplifiers: recent advances and applications*”, *Advances in Optics and Photonics*, vol. 14, no. 3, pp. 571-651, 2022.

⁷⁶ G. Giuliani and D. D’Alessandro, “*Noise analysis of conventional and gain-clamped semiconductor optical amplifiers*,” *Journal of Lightwave Technology*, vol. 18, pp. 1256-1263, 2000.

environment, remains common because it is transparent, modifiable, and naturally supports integration with DSP and learning pipelines⁷⁷.

Recent work on *end-to-end learned transceivers* makes this trend explicit: autoencoder-based PAM IM-DD designs treat the full link (including amplified noise) as a differentiable or simulation-driven channel during optimization, and then validate gains using conventional metrics⁷⁸. This direction is closely aligned with the “digital surrogate” philosophy: the simulator is not only a prediction tool, but also a *data generator* for training and optimization, meaning that model structure, parameterization, and logging are first-class requirements (not afterthoughts)⁷⁹.

3.3.1.5 From simulation to optimization loops (bridging toward surrogate-based design)

Finally, the literature increasingly treats modeling not as a one-shot evaluation, but as part of an *optimization loop*: tune operating points, fit compact parameters, and search trade-offs under constraints. Even before modern deep learning approaches, optical-component optimization studies demonstrated multi-objective search over device bias and configuration (e.g., SOA-assisted interferometric structures) using algorithmic optimization wrapped around models that predict gain, distortion, and noise-related penalties⁸⁰. In current IM-DD contexts, this naturally extends to surrogate-based workflows: (i) identify the minimal parameter set that reproduces measured behavior; (ii) calibrate; (iii) generate labeled datasets

⁷⁷ T. Wettlin; S. Calabrò; T. Rahman; J. Wei; N. Stojanovic, “*DSP for High-Speed Short-Reach IM/DD Systems Using PAM*”, Journal of Lightwave Technology, vol. 38, pp. 6771-6778, 2020.

⁷⁸ A. Omid, M. Banawan, E. Weckenmann, B. Paquin, A. Geravand, Z. Zheng, “*Auto-Encoder Optimized PAM IM/DD Transceivers for Amplified Fiber Links*”, Journal of Lightwave Technology, vol. 43, pp. 547-556, 2024.

⁷⁹ *Ibidem*.

⁸⁰ R. Dionisio, G. Parca, C. Reis, and A. Teixeira, “*Operational parameter optimisation of MZI-SOA using multi-objective genetic algorithms*,” Electronics Letters, vol. 47, pp. 561-562, 2011.

across conditions; and (iv) train ML models for correction, control, or joint TX/RX optimization⁸¹.

Modeling layer	Typical goal / use	Typical outputs	Calibration requirements
Analytical / semi-analytical (fast)	Link budgeting, quick trade studies, feasibility screening	OSNR/SNR targets; BER/EVM trends (approx.)	Datasheet/measurements for loss/gain, bandwidth, noise terms; optional fitted penalty
System-level waveform simulation (block-based time domain)	End-to-end feasibility; ISI/DSP interaction; evaluate PAM/BPAM and equalizers	BER/SER; EVM; eye metrics; spectra; OSNR $\leftrightarrow$ SNR mapping	Measured/assumed transfer functions; noise PSD/levels; DAC/ADC parameters; SOA parameters if used
Device-/circuit-aware co-simulation (highest fidelity)	Root-cause analysis; derive/validate compact models; predict new operating points	High-fidelity waveforms; distortion/memory characterization; fitted compact parameters	Detailed device/circuit measurements (curves, dynamics, S-params, parasitics); instrument-calibrated data

Table 3.1 Modeling and simulation stack for short-reach IM-DD links

3.3.2 Typical simplifications (ideal DAC/ADC, lumped noise, simplified SOA, etc.)

Even when the target application is the same (short-reach IM-DD links with PAM4/PAM8 and related formats), the modeling choices used in the literature and in common simulation workflows span a wide range of fidelity. In practice, many studies adopt a “minimum-complexity” model that is sufficient to compare modulation/DSP options under bandwidth and noise constraints, while treating hardware non-idealities as either idealized or absorbed into a small number of effective parameters⁸². This subsection highlights the most common simplifications, why they are attractive, and why they become limiting for a device-level digital surrogate intended to support ML training and optimization.

3.3.2.1 Idealized DAC/ADC interfaces and discrete-time abstraction

⁸¹ G. P. Agrawal, “Semiconductor optical amplifiers: recent advances and applications,” *Advances in Optics and Photonics*, vol. 14, no. 3, pp. 571-651, 2022.

⁸² X. Pang, *et al.* “200 Gb/s Per Lane IM/DD Technologies for Short Reach Optical Interconnects”, *Journal of Lightwave Technology*, vol. 38, pp. 492-503, 2020.

A frequent simplification is to represent the transmitter and receiver as *discrete-time* systems operating directly on sample sequences, with the DAC and ADC treated as transparent interfaces. In such models, the waveform entering the electrical driver is the ideal symbol-rate or oversampled sequence (often shaped by a pulse-shaping filter), and the receiver “observes” samples after an equivalent channel and additive noise, without explicitly modeling the DAC’s finite resolution, INL/DNL, zero-order hold (ZOH) behavior, glitch energy, sampling-clock jitter, or ADC quantization/clipping⁸³. This abstraction is particularly convenient for system-level comparisons and DSP algorithm design, because it yields a compact baseband model such as:

$$y[k] = (h * x)[k] + n[k],$$

where h aggregates bandwidth limitations and dispersion (if included), and $n[k]$ is a lumped noise term. It also aligns naturally with equalizer-centric studies, where the key question is how much ISI (i.e., effective channel memory) can be mitigated at a given oversampling ratio and DSP complexity⁸⁴.

However, for our objective, end-to-end optimization and a realistic surrogate, this simplification hides precisely the mechanisms that often dominate implementation penalties at high baud rates: waveform distortion introduced *before* the modulator (finite-resolution quantization, static DAC nonlinearity, ZOH spectral images, glitch-induced transients) and sampling/quantization effects introduced *after* the photodetection chain⁸⁵. As a result, the surrogate developed in this thesis intentionally preserves these interface details and increases the time resolution beyond what is typical in symbol-rate abstractions, so that analog-domain effects (e.g., post-ZOH waveform shape and glitch signatures) can be represented explicitly rather than “folded” into a generic noise term.

⁸³ T. Wettlin; S. Calabrò; T. Rahman; J. Wei; N. Stojanovic, “*DSP for High-Speed Short-Reach IM/DD Systems Using PAM*”, Journal of Lightwave Technology, vol. 38, pp. 6771-6778, 2020.

⁸⁴ Y. Yu, Y. Che, T. Bo, D. Kim, & H. Kim “*Reduced-state MLSE for an IM/DD system using PAM modulation*”, Optics Express, vol. 28, pp. 38505-38515, 2020.

⁸⁵ T. Wettlin; S. Calabrò; T. Rahman; J. Wei; N. Stojanovic, “*DSP for High-Speed Short-Reach IM/DD Systems Using PAM*”, Journal of Lightwave Technology, vol. 38, pp. 6771-6778, 2020.

3.3.2.2 Linear bandwidth limitation modeled as a single equivalent filter

Because bandwidth is a primary bottleneck in short-reach links, another common simplification is to collapse the electrical transmitter chain (driver plus modulator bandwidth) and the receiver chain (PD/TIA/AFE + ADC front-end) into one or two linear low-pass transfer functions⁸⁶. This is often justified: the dominant impairment in many PAM-based short-reach systems is linear ISI from bandwidth constraints, and the performance differences among equalization strategies (FFE/DFE, Tomlinson-Harashima precoding, decision-directed schemes, MLSE variants) can be studied effectively in a linear-filter framework⁸⁷. Even when bandwidth-extension techniques are the subject, the starting point is typically a linear time invariant (LTI) representation whose parameters are estimated from measured frequency responses or assumed templates⁸⁸.

The limitation is not the use of an LTI filter per se, but *where* that filter is placed and *what it replaces*. In low-fidelity studies, the entire mixed-signal/analog chain is reduced to a single $H(f)$ (overall end-to-end frequency response), and non-linear or data-dependent effects (e.g., DAC INL/DNL, SOA gain compression, PD/TIA saturation) are absent. For device-aware optimization, especially when the goal is to train ML models that must generalize across operating points, the absence of these nonlinearities is not a small detail; it can alter the statistics of the waveform and the error events that drive BER/EVM⁸⁹.

3.3.2.3 Lumped noise models and “effective SNR/OSNR” parameters

To keep analysis tractable, noise is frequently modeled as a single additive Gaussian process (electrical AWGN after detection) or a compact OSNR-to-BER

⁸⁶ X. Pang, *et al.* “200 Gb/s Per Lane IM/DD Technologies for Short Reach Optical Interconnects”, *Journal of Lightwave Technology*, vol. 38, pp. 492-503, 2020.

⁸⁷ T. Wettlin; S. Calabrò; T. Rahman; J. Wei; N. Stojanovic, “DSP for High-Speed Short-Reach IM/DD Systems Using PAM”, *Journal of Lightwave Technology*, vol. 38, pp. 6771-6778, 2020.

⁸⁸ C. Kottke, C. Schmidt, V. Jungnickel, & R. Freund, “Performance of Bandwidth Extension Techniques for High-Speed Short-Range IM/DD Links”, *Journal of Lightwave Technology*, vol. 37, pp. 665-672, 2019.

⁸⁹ X. Pang, *et al.* “200 Gb/s Per Lane IM/DD Technologies for Short Reach Optical Interconnects”, *Journal of Lightwave Technology*, vol. 38, pp. 492-503, 2020.

mapping when amplified links are considered⁹⁰. This approach enables fast sweeps and clean comparisons across modulation formats. It is also common in data-driven transceiver optimization, where the learning objective is defined over a simplified, differentiable channel model (e.g., AWGN plus bandwidth) to stabilize training and make it computationally feasible⁹¹.

From a surrogate perspective, the issue is that real IM-DD receivers rarely encounter a single white-noise source. Shot noise, thermal noise, relative intensity noise (RIN), TIA noise, quantization noise, and (if present) optical amplifier ASE contribute with different dependencies on optical power, bandwidth, and operating point. When these are collapsed into a single effective variance, the model may predict the “right” BER under one condition but fail to extrapolate when bias current, received power, temperature, or bandwidth change, precisely the kind of extrapolation needed for optimization and robust ML training⁹².

3.3.2.4 Simplified SOA modeling (static gain + ASE) and limited dynamics

When SOAs appear in IM-DD studies (e.g., reach extension, PON/fronthaul sensitivity improvement), a widely used simplification is to treat the SOA as: (i) a constant gain block, sometimes with a basic gain-compression law; plus (ii) an ASE noise source characterized by a noise figure, with limited or no dynamic carrier-recovery effects⁹³. For many engineering comparisons, this is reasonable: it captures first-order sensitivity improvements and OSNR penalties without requiring stiff carrier-rate equations.

Yet SOAs are inherently nonlinear, with memory effects tied to carrier dynamics and intraband processes. Detailed treatments show how gain saturation,

⁹⁰ X. Zhou, R. Urata, and H. Liu, “Beyond 1 Tb/s Intra-Data Center Interconnect Technology: IM-DD OR Coherent?”, *Journal of Lightwave Technology*, vol. 38, pp. 475-484, 2020.

⁹¹ A. Omid, M. Banawan, E. Weckenmann, B. Paquin, A. Geravand, Z. Zheng, “Auto-Encoder Optimized PAM IM/DD Transceivers for Amplified Fiber Links”, *Journal of Lightwave Technology*, vol. 43, pp. 547-556, 2024.

⁹² *Ibidem*.

⁹³ J. Zhang, *et al.* “SOA Pre-Amplified 100 Gb/s/λ PAM-4 TDM-PON Downstream Transmission Using 10 Gbps O-Band Transmitters”, *Journal of Lightwave Technology*, vol. 38, pp. 185-193, 2020.

recovery time, and amplitude-phase coupling can reshape multi-level waveforms and introduce pattern dependence, effects that may matter for high-baud PAM and become even more pronounced when advanced DSP relies on waveform structure or reconstruction mechanisms⁹⁴. This motivates a higher-fidelity SOA block in the surrogate than the “gain plus ASE” approximation, even if the thesis later uses reduced models for specific sweeps.

3.3.2.5 Fiber channel is often treated as “nearly transparent” in short-reach regimes

In intra-data-center and very short reach, it is common to neglect fiber nonlinearities and include dispersion only when the reach and wavelength band make it unavoidable. Many short-reach studies assume that the dominant limitations are transceiver bandwidth and noise, whereas the fiber contributes primarily to attenuation (and sometimes to a simple dispersive filter)⁹⁵. This is consistent with the practical reality that many O-band links are engineered to minimize chromatic dispersion penalties over relevant reaches.

For an end-to-end surrogate, however, the key point is not that fiber effects are always large, but that the *model should be able to account for them consistently* as the operating regime changes (reach extension, different band, tighter spectral packing, or interactions with SOA chirp and filtering). A modular surrogate benefits from maintaining a unified time-domain waveform representation that can incorporate dispersion, filtering, and device-induced chirp within the same signal path when required⁹⁶.

3.3.2.6 Limited time resolution and omission of “analog waveform realism.”

⁹⁴ A. Sobhanan, A. Anthur, S. O’Duill, M. Pelusi, S. Namiki, L. Barry, D. Venkitesh, and G. P Agrawal, “Semiconductor optical amplifiers: recent advances and applications”, Advances in Optics and Photonics, vol. 14, no. 3, pp. 571-651, 2022.

⁹⁵ X. Zhou, R. Urata, and H. Liu, “Beyond 1 Tb/s Intra-Data Center Interconnect Technology: IM-DD OR Coherent?”, Journal of Lightwave Technology, vol. 38, pp. 475-484, 2020.

⁹⁶ A. Sobhanan, A. Anthur, S. O’Duill, M. Pelusi, S. Namiki, L. Barry, D. Venkitesh, and G. P Agrawal, “Semiconductor optical amplifiers: recent advances and applications”, Advances in Optics and Photonics, vol. 14, no. 3, pp. 571-651, 2022.

Finally, many simulations use symbol-rate or modest oversampling representations that are sufficient for conventional PAM detection and equalization, without explicitly resolving the fine-time structure introduced by ZOH behavior, glitch transients, or high-frequency spectral images⁹⁷. This choice is sensible for throughput and for studies focused on DSP/architecture trade-offs. However, this limitation arises when the research goal shifts toward *device-level impairment learning* and *hardware-aware optimization*, in which the model should expose waveform signatures that ML or advanced DSP can exploit (or must robustly tolerate).

This is also directly relevant when moving beyond conventional unipolar PAM toward bipolar reconstruction schemes: if the receiver processing depends on oversampling and field/sign reconstruction, then waveform-shaping impairments with memory (including SOA dynamics, mixed-signal artifacts, or receiver front-end bandwidth interactions) can perturb the structure on which reconstruction relies⁹⁸. Compared with models that operate purely at symbol rate, the proposed surrogate offers the advantage of representing the analog waveform after digital-to-analog conversion with finer time granularity, including effects such as zero-order hold and transient glitching. This level of detail is essential when the objective is to understand how hardware-induced distortions propagate through the IM/DD chain and to construct simulation datasets that reflect the signal structures an AI model would ultimately need to learn from in practical optimization workflows.

3.4 Digital twins and digital surrogates in communications and photonics

3.4.1 Concepts and definitions

The term digital twin is now used widely across engineering disciplines, but its meaning depends on *how tightly* the virtual model is coupled to the physical system.

⁹⁷ T. Wettlin; S. Calabrò; T. Rahman; J. Wei; N. Stojanovic, “*DSP for High-Speed Short-Reach IM/DD Systems Using PAM*”, Journal of Lightwave Technology, vol. 38, pp. 6771-6778, 2020.

⁹⁸ M. Secondini and E. Forestieri, “*Direct Detection of Bipolar Pulse amplitude modulation*”, Journal of Lightwave Technology, 2020.

Systematic reviews generally trace the origin of the concept to early product-lifecycle discussions (often attributed to M. Grieves and collaborators, and later adopted and popularized in aerospace)⁹⁹, and emphasize that what distinguishes a digital twin from ordinary simulation is the presence of an explicit *connection mechanism* that keeps the virtual representation aligned with the evolving state of its physical counterpart¹⁰⁰.

One method of formalizing this terminology is through the widely recognized continuum of *digital model*, *digital shadow*, and *digital twin*. In this classification, a *digital model* is an offline representation that is not automatically synchronized; a *digital shadow* introduces *automatic, typically one-way* data flow from the physical system to the digital instance (the model is updated as measurements arrive); and a *digital twin* implies a tighter coupling where the digital instance can also influence the physical system (i.e., *bi-directional* interaction enabling monitoring *and* decision/control loops)¹⁰¹. This continuum clarifies why different communities sometimes disagree: two works may both use the term “digital twin,” while one is effectively an offline simulator and the other is a synchronized, operational replica.

Beyond synchronization, the literature also converges on a set of *structural elements* that appear repeatedly in definitions: (i) a *physical entity* (asset/system/process), (ii) a *digital representation* (physics-based, data-driven, or hybrid), and (iii) *data and service interfaces* (sensing/telemetry, model updating,

⁹⁹ E. H. Glaessgen and D. S. Stargel, “*The Digital Twin Paradigm for Future NASA and U.S. Air Force Vehicles*”, AIAA Paper 2012-1818, 2012.

¹⁰⁰ D. Jones, C. Snider, A. Nassehi, J. Yon, and B. Hicks, “*Characterising the Digital Twin: A systematic literature review*”, CIRP Journal of Manufacturing Science and Technology, vol. 29, pp. 36-52, 2020.

¹⁰¹ W. Kritzinger, M. Karner, G. Traar, J. Henjes, and W. Sihn, “*Digital Twin in manufacturing: A categorical literature review and classification*”, IFAC-PapersOnLine, vol. 51, no. 11, pp. 1016-1022, 2018.

prediction, and decision support)^{102,103}. In practice, the “digital representation” may be multi-resolution (e.g., combining first-principles blocks with empirical sub-models), while the interface layer defines how calibration, state estimation, and uncertainty are handled. Such issues become central once a twin is used for autonomous operation rather than for offline analysis¹⁰⁴.

Within this thesis, the term *digital surrogate* is used in a more *model-centric* sense: a surrogate is a digital proxy constructed to reproduce selected input-output behaviors of a system with sufficient accuracy for a given task (e.g., sensitivity prediction, BER/EVM estimation, control/optimization), without requiring that the model be continuously synchronized to a specific deployed asset. In other words, a surrogate can be viewed as a *functionally faithful emulator* for experimentation and optimization. In contrast, a digital twin is best understood as a *surrogate plus connectivity plus lifecycle integration* (calibration/updates and, in mature forms, operational feedback)^{105,106}. Under this perspective, the duties of the two notions overlap substantially whenever the surrogate is (a) calibrated against measurements and (b) periodically or continuously updated. In that regime, “digital surrogate” and “digital twin” can become interchangeable *in practice*, even if the stricter, bidirectional definition of a twin is not fully satisfied.

This definitional ambiguity is visible in optical communications, where “digital twin” is increasingly used to describe (1) *network-level* replicas enabling

¹⁰² D. Jones, C. Snider, A. Nassehi, J. Yon, and B. Hicks, “*Characterising the Digital Twin: A systematic literature review*”, CIRP Journal of Manufacturing Science and Technology, vol. 29, pp. 36-52, 2020.

¹⁰³ B. R. Barricelli, E. Casiraghi, and D. Fogli, “*A survey on digital twin: Definitions, characteristics, applications, and design implications*”, IEEE Access, vol. 7, pp. 167653-167671, 2019.

¹⁰⁴ C. Semeraro, M. Lezoche, H. Panetto, and M. Dassisti, “*Digital twin paradigm: A systematic literature review*”, Computers in Industry, vol. 130, 103469, 2021.

¹⁰⁵ W. Kritzinger, M. Karner, G. Traar, J. Henjes, and W. Sihn, “*Digital Twin in manufacturing: A categorical literature review and classification*”, IFAC-PapersOnLine, vol. 51, no. 11, pp. 1016-1022, 2018.

¹⁰⁶ *Ibidem* 104. (C. Semeraro *et al.*)

automation¹⁰⁷, and (2) *device-/instrument-level* replicas enabling advanced monitoring and inference¹⁰⁸. In network operations, digital-twin workflows are typically motivated by fault management, configuration validation, and transmission-performance prediction, in which a digital representation of the optical layer is connected to operational data to accelerate troubleshooting and enable closed-loop control^{109,110}. At the instrument level, the “twin” may be a data-driven model that maps measured waveforms to latent physical parameters or impairment fingerprints; for example, deep-learning-based twins have been proposed for time-domain optical analyzers to infer physical events and accelerate analysis from measurement traces¹¹¹. These examples show why the term is attractive in photonics: the combination of (i) complex physics, (ii) incomplete observability, and (iii) high operational cost of experimentation makes calibrated digital replicas valuable even when they act primarily as predictive surrogates rather than fully bidirectional twins.

For IM/DD optical transmission and end-to-end research, these distinctions matter because they shape what can be claimed and what can be reliably optimized. If a model is used mainly for offline “what-if” evaluation, it is closer to a *surrogate* or *digital model*. If it is continuously corrected using lab/field measurements (component responses, noise statistics, drift, aging), it moves toward a *digital shadow*. If it also supports closed-loop actions (e.g., suggesting and deploying

¹⁰⁷ R. Vilalta; L. Gifre; R. Casellas; R. Muñoz; R. Martínez; A. Mozo, “Applying Digital Twins to Optical Networks with Cloud-Native SDN Controllers”, IEEE Communications Magazine, pp. 129-133, 2023.

¹⁰⁸ D. Wang, Z. Zhang, M. Zhang, M. Fu, J. Li, S. Cai “The Role of Digital Twin in Optical Communication: Fault Management, Hardware Configuration, and Transmission Simulation”, IEEE Communications Magazine, pp. 133-139, 2021.

¹⁰⁹ *Ibidem*.

¹¹⁰ X. Liu, Y. Zhang, Y. Chen, Y. Liu, M. Cai, Q. Qiu, M. Fu, L. Yi, W. Hu, Q. Zhuge, “Digital twin modeling and controlling of optical power evolution enabling autonomous-driving optical networks: a Bayesian approach”, Advanced Photonics, vol. 6, no. 2, 026006, 2024.

¹¹¹ D. Sequeira, M. Ruiz, N. Costa, A. Napoli, J. P. Pedro, and L. Velasco, “OCATA: a deep-learning-based digital twin for the optical time domain analyzer”, Journal of Optical Communications and Networking, vol. 15, no. 2, pp. 87-97, 2023.

parameter changes at the transmitter/receiver or network controller), it approaches the stronger notion of a *digital twin*^{112,113,114}. Importantly, ML-driven optimization benefits from both: surrogates provide scalable, controllable training environments, while twin-like coupling provides *reality anchoring* through systematic calibration and update, reducing the risk that learned policies exploit simulation artifacts rather than physical behavior^{115,116}.

3.4.2 Typical scope (network-level vs device-level)

In practice, “digital twin” work in optical communications spans multiple abstraction layers. The scope is primarily determined by (i) *what the “asset” is* (network, node, link, component, instrument), (ii) *what is synchronized* (configuration/state, performance metrics, or full physics), and (iii) *the time scale* at which the virtual model must track the physical system. A useful way to frame this is as a continuum from *network-level twins* (topology/control/state and service performance) down to *device-level twins* (component physics and waveform- or sample-level behavior), with “hybrid” and hierarchical approaches in between¹¹⁷.

¹¹² W. Kritzing, M. Karner, G. Traar, J. Henjes, and W. Sih, “*Digital Twin in manufacturing: A categorical literature review and classification*”, IFAC-PapersOnLine, vol. 51, no. 11, pp. 1016-1022, 2018.

¹¹³ D. Wang, Z. Zhang, M. Zhang, M. Fu, J. Li, S. Cai “*The Role of Digital Twin in Optical Communication: Fault Management, Hardware Configuration, and Transmission Simulation*”, IEEE Communications Magazine, pp. 133-139, 2021.

¹¹⁴ R. Vilalta; L. Gifre; R. Casellas; R. Muñoz; R. Martínez; A. Mozo, “*Applying Digital Twins to Optical Networks with Cloud-Native SDN Controllers*”, IEEE Communications Magazine, pp. 129-133, 2023.

¹¹⁵ C. Semeraro, M. Lezoche, H. Panetto, and M. Dassisti, “*Digital twin paradigm: A systematic literature review*”, Computers in Industry, vol. 130, 103469, 2021.

¹¹⁶ X. Liu, Y. Zhang, Y. Chen, Y. Liu, M. Cai, Q. Qiu, M. Fu, L. Yi, W. Hu, Q. Zhuge, “*Digital twin modeling and controlling of optical power evolution enabling autonomous-driving optical networks: a Bayesian approach*”, Advanced Photonics, vol. 6, no. 2, 026006, 2024.

¹¹⁷ D. Jones, C. Snider, A. Nassehi, J. Yon, and B. Hicks, “*Characterising the Digital Twin: A systematic literature review*”, CIRP Journal of Manufacturing Science and Technology, vol. 29, pp. 36-52, 2020.

3.4.2.1 Network-level digital twins (operations, control, and service/QoT intent).

A network-level digital twin typically represents the optical network as a *virtual replica of the infrastructure and its operational state*, including topology, resource occupancy (wavelengths/spectrum, routes), configuration parameters (e.g., power setpoints, channel plans), alarms, and telemetry-derived health indicators. Its primary purpose is *network decision support and automation*: planning, what-if analysis, closed-loop control, fault localization, and service assurance¹¹⁸. In this scope, the “twin” is often tightly coupled to software-defined networking (SDN)/orchestration logic: it may run as a cloud-native service and interact with an SDN controller to keep the virtual state consistent with the deployed network and to test control actions before (or while) applying them to the physical system¹¹⁹.

Crucially, in many network-level twins the physical layer is not modeled at waveform fidelity. Instead, it is represented through *engineering abstractions* (e.g., QoT predictors, margin models, or semi-analytical performance estimators) that are computationally light enough for frequent evaluation across many candidate configurations¹²⁰. This is not a weakness per se; it is a design choice matching the *time scales* of network operations (seconds to hours) and the need to explore large configuration spaces. The common output is therefore *network-facing*: feasible provisioning decisions, predicted quality of transmission (QoT) penalties, expected OSNR/SNR margins at the service level, and fault-management actions^{121,122}.

¹¹⁸ D. Wang, Z. Zhang, M. Zhang, M. Fu, J. Li, S. Cai “*The Role of Digital Twin in Optical Communication: Fault Management, Hardware Configuration, and Transmission Simulation*”, IEEE Communications Magazine, pp. 133-139, 2021.

¹¹⁹ R. Vilalta; L. Gifre; R. Casellas; R. Muñoz; R. Martínez; A. Mozo, “*Applying Digital Twins to Optical Networks with Cloud-Native SDN Controllers*”, IEEE Communications Magazine, pp. 129-133, 2023.

¹²⁰ *Ibidem*.

¹²¹ R. Vilalta; L. Gifre; R. Casellas; R. Muñoz; R. Martínez; A. Mozo, “*Applying Digital Twins to Optical Networks with Cloud-Native SDN Controllers*”, IEEE Communications Magazine, pp. 129-133, 2023.

¹²² D. Wang, Z. Zhang, M. Zhang, M. Fu, J. Li, S. Cai “*The Role of Digital Twin in Optical Communication: Fault Management, Hardware Configuration, and Transmission Simulation*”, IEEE Communications Magazine

3.4.2.2 Device-/link-level digital twins (physical-layer behavior and component dynamics).

At the device/link level, the “asset” is a *specific physical subsystem* (e.g., an optical amplifier chain, a transceiver, a fiber span/link segment, or a measurement instrument). Here, the twins’ value comes from representing *component dynamics and physics-constrained behavior* that directly shape signal quality. For example, a device-oriented twin can be used to infer and control *optical power evolution* in an autonomous network setting, using a probabilistic (Bayesian) model that continuously incorporates observations, enabling control decisions explicitly tied to physical-layer evolution rather than to abstract QoT margins¹²³.

A second important device-level use case is when the twin emulates a measurement instrument (or its measurement process) to accelerate monitoring and diagnosis. OCATA is an example in which a deep-learning-based twin is built for an optical time-domain analyzer workflow, aiming to reproduce instrument-level outcomes with a learned model, thereby shifting part of the monitoring/interpretation burden from repeated physical measurements to a calibrated virtual replica¹²⁴. In these device-/instrument-level scopes, the twin is typically calibrated with lab/field data and is evaluated on *physical-layer outputs*: predicted power profiles, gain/tilt evolution, instrument traces, impairment estimates, or component-state inference, and often at finer temporal or spatial granularity than network-level controllers require¹²⁵.

3.4.2.3 Hybrid and hierarchical scope (composable twins across layers).

Many realistic deployments end up being *hierarchical*: a network-level twin orchestrates decisions and maintains configuration/state, while embedding *lower-level surrogate/twin models* for the physical layer (component/link behavior) where

¹²³ X. Liu, W. Hu, K. Zhang, A. Xu, Q. Zhou, K. Wang, and J. Yu, “Digital twin modeling and controlling of optical power evolution enabling autonomous-driving optical networks: a Bayesian approach”, *Advanced Photonics*, vol. 6, no. 2, 026006, 2024.

¹²⁴ D. Sequeira, M. Ruiz, N. Costa, A. Napoli, J. P. Pedro, and L. Velasco, “OCATA: a deep-learning-based digital twin for the optical time domain analyzer”, *Journal of Optical Communications and Networking*, vol. 15, no. 2, pp. 87-97, 2023.

¹²⁵ *Ibidem*.

higher fidelity is needed¹²⁶. This hierarchy is where terminology and “duties” can overlap: a network digital twin may contain (or call) device-level models that serve as *digital surrogates* for subsystems. At the same time, the top layer maintains the “twin” characteristics of state synchronization and operational alignment with the physical network¹²⁷. The key engineering trade-off is explicit: more physics and higher time resolution improve explanatory power and calibration fidelity, but increase computational cost and the amount of measurement data required to maintain model validity across operating conditions¹²⁸.

The digital surrogate developed in this thesis is aligned with the *device-/link-level* end of this spectrum: it targets the end-to-end physical-layer behavior of the IM-DD chain (including mixed-signal and component non-idealities) to support waveform-realistic performance evaluation and the generation of structured data suitable for training ML models. Conceptually, such a surrogate can also serve as a *physical-layer module* within a broader network-level digital twin (e.g., as a high-fidelity “plug-in” for specific links, transceivers, or amplifier chains). Still, its primary scope remains the detailed representation of subsystems required for end-to-end optimization studies.

3.4.3 Limitations (partial models, not end-to-end, limited calibration)

Despite the strong momentum around digital twins and closely related “digital surrogate” approaches, the literature also highlights recurring limitations that are especially relevant in optical communications, where performance depends on tightly coupled physical-layer effects, device operating points, and mixed-signal impairments¹²⁹. A first limitation is scope fragmentation: many so-called twins are built for a specific operational task (e.g., fault management, configuration support,

¹²⁶ D. Jones, C. Snider, A. Nassehi, J. Yon, and B. Hicks, “*Characterising the Digital Twin: A systematic literature review*”, CIRP Journal of Manufacturing Science and Technology, vol. 29, pp. 36-52, 2020.

¹²⁷ *Ibidem*.

¹²⁸ D. Wang, Z. Zhang, M. Zhang, M. Fu, J. Li, S. Cai “*The Role of Digital Twin in Optical Communication: Fault Management, Hardware Configuration, and Transmission Simulation*”, IEEE Communications Magazine, pp. 133-139, 2021.

¹²⁹ D. Jones, C. Snider, A. Nassehi, J. Yon, and B. Hicks, “*Characterising the Digital Twin: A systematic literature review*”, CIRP Journal of Manufacturing Science and Technology, vol. 29, pp. 36-52, 2020.

QoT estimation, or localized prediction of a network metric), rather than as a unified representation of the full transmission chain¹³⁰. Systematic reviews emphasize that the term “digital twin” is used inconsistently across communities, sometimes denoting little more than an offline simulation model and data visualization, and sometimes a live, synchronized model with bidirectional interaction¹³¹. This ambiguity makes it difficult to compare approaches rigorously and can lead to “partial” twins being interpreted as comprehensive representations.

A second limitation is the fidelity-complexity trade-off, which is particularly acute in communications. In optical networking and transmission, a high-fidelity physical model (e.g., detailed propagation or device-level dynamics) is computationally expensive and often unsuitable for real-time operation or large-scale parameter sweeps, motivating reduced-order, semi-empirical, or ML-based surrogates^{132,133}. However, reducing complexity typically requires choosing which effects to preserve and which to ignore, and these choices are often application-driven: a twin optimized for QoT estimation may not reproduce waveform- or constellation-level behavior in a way that supports impairment diagnosis or end-to-end optimization¹³⁴. For this reason, many published “twins” remain limited to a subset of phenomena (e.g., amplifier noise and filtering in network scenarios), and

¹³⁰ M. M. Rathore, S. A. Shah, Dh. Shukla, E. Bentaftat, S. BAKIRAS, “*The Role of AI, Machine Learning, and Big Data in Digital Twinning: A Systematic Literature Review, Challenges, and Opportunities*”, IEEE Access, vol. 9, pp. 32030-32052, 2021.

¹³¹ C. Semeraro, M. Lezoche, H. Panetto, and M. Dassisti, “*Digital twin paradigm: A systematic literature review*”, Computers in Industry, vol. 130, 103469, 2021.

¹³² T. Kreuzer, P. Papapetrou, J. Zdravkovic, “*Artificial intelligence in digital twins - A systematic literature review*”, Data & Knowledge Engineering, vol. 151, art. 102304, 2024.

¹³³ R. Vilalta; L. Gifre; R. Casellas; R. Muñoz; R. Martínez; A. Mozo, “*Applying Digital Twins to Optical Networks with Cloud-Native SDN Controllers*”, IEEE Communications Magazine, pp. 129-133, 2023.

¹³⁴ D. Sequeira, M. Ruiz, N. Costa, A. Napoli, J. P. Pedro, and L. Velasco, “*OCATA: a deep-learning-based digital twin for the optical time domain analyzer*”, Journal of Optical Communications and Networking, vol. 15, no. 2, pp. 87-97, 2023.

they do not jointly model mixed-signal waveform formation, analog front-end behavior, and component nonlinearities in a unified manner¹³⁵.

A third limitation is *calibration burden and observability*. A twin that aims to track a real system requires reliable access to measurements and, ideally, to internal state variables. Yet, optical systems expose only a limited set of observables (e.g., received power, OSNR estimates, monitored spectra/constellations, alarms, or counters). These measurements may be vendor-dependent or partially hidden in proprietary domains¹³⁶. Even when measurement access exists, calibration is not a one-time task: device characteristics drift with temperature, aging, configuration changes, and network re-routing^{137,138}. Reviews of digital twin practice repeatedly note that calibration quality and the availability of trustworthy data streams often become the dominant practical constraint, more than the modeling method itself¹³⁹. In optical networks, this challenge is explicit in work that differentiates fully disaggregated scenarios (where component information and telemetry are accessible) from mixed disaggregated proprietary scenarios (where the internal configuration of vendor segments is not observable, so the twin must infer behavior from end-point measurements)¹⁴⁰.

A related limitation is *generalization and domain shift*. Data-driven twins can achieve excellent accuracy within their training envelope, but their reliability may degrade when operating conditions depart from that envelope (e.g., new traffic

¹³⁵ D. Sequeira, M. Ruiz, N. Costa, A. Napoli, J. P. Pedro, and L. Velasco, “OCATA: a deep-learning-based digital twin for the optical time domain analyzer”, *Journal of Optical Communications and Networking*, vol. 15, no. 2, pp. 87-97, 2023.

¹³⁶ *Ibidem*.

¹³⁷ M. M. Rathore, H. Bin Khalifa, S. A. Shah, Dh. Shukla, E. Bentafat, S. BAKIRAS, “The Role of AI, Machine Learning, and Big Data in Digital Twinning: A Systematic Literature Review, Challenges, and Opportunities”, *IEEE Access*, vol. 9, pp. 32030-32052, 2021.

¹³⁸ R. Vilalta; L. Gifre; R. Casellas; R. Muñoz; R. Martínez; A. Mozo, “Applying Digital Twins to Optical Networks with Cloud-Native SDN Controllers”, *IEEE Communications Magazine*, pp. 129-133, 2023.

¹³⁹ D. Jones, C. Snider, A. Nassehi, J. Yon, and B. Hicks, “Characterising the Digital Twin: A systematic literature review”, *CIRP Journal of Manufacturing Science and Technology*, vol. 29, pp. 36-52, 2020.

¹⁴⁰ *Ibidem* 138.(R. Vilalta *et al.*)

mixes, power profiles, channel counts, component variants, or unmodeled impairments)¹⁴¹. This is especially important when the twin is used for decision-making (e.g., autonomous control, configuration optimization, or fault localization), because extrapolation errors can lead to unstable or incorrect actions. As a consequence, many frameworks emphasize either careful scoping (restricting the twin to well-defined operating regimes) or the need for continuous adaptation and retraining, both of which increase engineering overhead^{142,143}.

A fifth limitation is the *need for uncertainty awareness and validation*. Optical environments contain intrinsic stochasticity (ASE, nonlinear noise-like effects, measurement noise), and practical twins must distinguish between modeling error and natural variability¹⁴⁴. Recent work on optical networks, therefore, explores probabilistic approaches and uncertainty-aware modeling to support safer decision-making and more robust autonomous control¹⁴⁵. Nonetheless, the validation problem remains non-trivial: demonstrating that a twin is accurate for a given task (and remains accurate over time) requires a systematic validation protocol, coverage metrics, and acceptance thresholds, elements that are often under-

¹⁴¹ *Ibidem*.

¹⁴² M. M. Rathore, H. Bin Khalifa, S. A. Shah, Dh. Shukla, E. Bentafat, S. BAKIRAS, “*The Role of AI, Machine Learning, and Big Data in Digital Twinning: A Systematic Literature Review, Challenges, and Opportunities*”, IEEE Access, vol. 9, pp. 32030-32052, 2021.

¹⁴³ T. Kreuzer, P. Papapetrou, J. Zdravkovic, “Artificial intelligence in digital twins - A systematic literature review”, Data & Knowledge Engineering, vol. 151, art. 102304, 2024.

¹⁴⁴ R. Vilalta; L. Gifre; R. Casellas; R. Muñoz; R. Martínez; A. Mozo, “*Applying Digital Twins to Optical Networks with Cloud-Native SDN Controllers*”, IEEE Communications Magazine, pp. 129-133, 2023.

¹⁴⁵ X. Liu, W. Hu, K. Zhang, A. Xu, Q. Zhou, K. Wang, and J. Yu, “*Digital twin modeling and controlling of optical power evolution enabling autonomous-driving optical networks: a Bayesian approach*”, Advanced Photonics, vol. 6, no. 2, 026006, 2024.

specified in application papers and remain an open practical issue across digital-twin domains^{146,147}.

Finally, the literature points to *governance and integration constraints*. Deploying a twin in a live communications environment raises issues beyond modeling accuracy: data access permissions, reproducibility, vendor interoperability, and the ability to integrate with software-defined networking (SDN)/Network management system (NMS) stacks without relying on proprietary assumptions^{148,149}. These constraints can implicitly push implementations toward partial models that rely solely on readily available telemetry, even when a richer, physics-grounded model would be desirable.

Taken together, these limitations motivate a careful positioning of any “twin/surrogate” claim: what is represented, at what level of fidelity, how it is calibrated, what telemetry it assumes, and which outputs it is intended to reproduce (metrics vs signals vs end-to-end behavior)^{150,151}. In the context of IM/DD research, where mixed-signal details and component operating points can materially affect multilevel formats and DSP/ML behavior, a recurring gap in the broader digital-twin literature is that many implementations remain *neither end-to-end* nor

¹⁴⁶ D. Jones, C. Snider, A. Nassehi, J. Yon, and B. Hicks, “*Characterising the Digital Twin: A systematic literature review*”, CIRP Journal of Manufacturing Science and Technology, vol. 29, pp. 36-52, 2020.

¹⁴⁷ C. Semeraro, M. Lezoche, H. Panetto, and M. Dassisti, “*Digital twin paradigm: A systematic literature review*”, Computers in Industry, vol. 130, 103469, 2021.

¹⁴⁸ D. Wang, Z. Zhang, M. Zhang, M. Fu, J. Li, S. Cai “*The Role of Digital Twin in Optical Communication: Fault Management, Hardware Configuration, and Transmission Simulation*”, IEEE Communications Magazine, pp. 133-139, 2021.

¹⁴⁹ D. Sequeira, M. Ruiz, N. Costa, A. Napoli, J. P. Pedro, and L. Velasco, “*OCATA: a deep-learning-based digital twin for the optical time domain analyzer*”, Journal of Optical Communications and Networking, vol. 15, no. 2, pp. 87-97, 2023.

¹⁵⁰ D. Jones, C. Snider, A. Nassehi, J. Yon, and B. Hicks, “*Characterising the Digital Twin: A systematic literature review*”, CIRP Journal of Manufacturing Science and Technology, vol. 29, pp. 36-52, 2020.

¹⁵¹ M. M. Rathore, H. Bin Khalifa, S. A. Shah, Dh. Shukla, E. Bentafat, S. BAKIRAS, “*The Role of AI, Machine Learning, and Big Data in Digital Twinning: A Systematic Literature Review, Challenges, and Opportunities*”, IEEE Access, vol. 9, pp. 32030-32052, 2021.

waveform-faithful beyond their specific target use case¹⁵²¹⁵³. This observation directly motivates the thesis direction toward a modular, calibratable surrogate that explicitly manages the fidelity-complexity trade-off at the component and waveform level, while remaining compatible with data-driven optimization workflows.

3.5 DSP and machine learning in optical communications

3.5.1 Classical DSP techniques for IM/DD PAM/BPAM

In short-reach IM/DD links, the dominant performance limitations are often rooted in bandwidth constraints and mixed-signal impairments rather than purely optical propagation effects. As a result, receiver- and transmitter-side digital signal processing (DSP) has become a central “workhorse” for enabling high line rates with multilevel PAM under practical electrical/optical hardware limits¹⁵⁴. In this context, “classical DSP” refers to algorithmic, model-driven methods (typically linear or weakly nonlinear) that are designed from communication theory and signal processing principles and are implemented with relatively transparent parameters (filter taps, decision rules, update laws), in contrast to fully data-driven models that learn end-to-end mappings (treated in Section 3.5.2). Classical DSP remains the baseline for today’s standardized PAM4 links and serves as a necessary reference point for emerging formats, including BPAM, because it defines the core equalization, synchronization, and detection functions that any “surrogate-aware” optimization must ultimately interface with¹⁵⁵.

¹⁵² R. Vilalta; L. Gifre; R. Casellas; R. Muñoz; R. Martínez; A. Mozo, “*Applying Digital Twins to Optical Networks with Cloud-Native SDN Controllers*”, IEEE Communications Magazine, pp. 129-133, 2023.

¹⁵³ D. Sequeira, M. Ruiz, N. Costa, A. Napoli, J. P. Pedro, and L. Velasco, “*OCATA: a deep-learning-based digital twin for the optical time domain analyzer*”, Journal of Optical Communications and Networking, vol. 15, no. 2, pp. 87-97, 2023.

¹⁵⁴ Xi. Pang, O. Ozolins, R. Lin, L. Zhang, A. Udalcovs, L. Xue, R. Schatz, U. Westergren, S. Xiao, W. Hu, G. Jacobsen, S. Popov, & J. Chen, “*200 Gb/s Per Lane IM/DD Technologies for Short Reach Optical Interconnects*”, Journal of Lightwave Technology, vol. 38, pp. 492–503, 2020.

¹⁵⁵ T. Wettlin, S. Calabrò, T. Rahman, J. Wei, and N. Stojanovic, “*DSP for High-Speed Short-Reach IM/DD Systems Using PAM*”, Journal of Lightwave Technology, vol. 38, pp. 6771–6778, 2020.

At a high level, the digital receiver chain for PAM/BPAM typically includes: (i) front-end conditioning (resampling, timing recovery), (ii) equalization (linear and/or nonlinear), (iii) detection and possible sequence estimation, and (iv) post-processing for performance monitoring (e.g., EVM/BER estimation, eye metrics). The specific choices depend on the severity of intersymbol interference (ISI), the extent of nonlinear distortions (DAC/driver/modulator, SOA, PD/TIA/ADC chain), and the complexity and latency constraints of the target interconnect¹⁵⁶.

3.5.1.1 Linear equalization (FFE/DFE) as the baseline.

The most common classical mitigation for bandwidth-limited IM/DD links is linear equalization, typically implemented as a feed-forward equalizer (FFE), i.e., a finite-impulse-response (FIR) filter applied to the sampled waveform before symbol decisions¹⁵⁷. In practice, FFEs are often combined with a decision-feedback equalizer (DFE), where previously detected symbols are fed back through another FIR filter to cancel post-cursor ISI. FFE/DFE structures are attractive because they map well onto hardware, can be adapted online, and provide a strong first-order correction for the “low-pass” nature of typical short-reach transmitter/receiver chains. Survey work on PAM4 equalization highlights that many practical performance differences among receiver designs arise from how effectively linear equalization manages ISI while maintaining stability in the presence of noise and decision errors¹⁵⁸.

Linear equalizers are often adapted using standard least-mean-square (LMS) or recursive least-squares (RLS) updates in training-based or decision-directed modes¹⁵⁹. Their effectiveness, however, is fundamentally limited when the channel exhibits strong nonlinearities or significant memory effects that a linear filter cannot

¹⁵⁶ Xi. Pang, O. Ozolins, R. Lin, L. Zhang, A. Udalcovs, L. Xue, R. Schatz, U. Westergren, S. Xiao, W. Hu, G. Jacobsen, S. Popov, & J. Chen, “200 Gb/s Per Lane IM/DD Technologies for Short Reach Optical Interconnects”, *Journal of Lightwave Technology*, vol. 38, pp. 492–503, 2020.

¹⁵⁷ T. Wettlin, S. Calabrò, T. Rahman, J. Wei, and N. Stojanovic, “DSP for High-Speed Short-Reach IM/DD Systems Using PAM”, *Journal of Lightwave Technology*, vol. 38, pp. 6771–6778, 2020.

¹⁵⁸ *Ibidem*.

¹⁵⁹ *Ibidem*.

capture. This limitation becomes more pronounced as symbol rates approach the bandwidth limits of DAC/driver/modulator and PD/TIA/ADC front-ends, where the residual distortion after linear equalization can dominate EVM and pre-FEC BER¹⁶⁰.

3.5.1.2 Transmitter-side pre-compensation: pre-emphasis and precoding.

To relieve receiver complexity or to compensate for deterministic bandwidth roll-off, short-reach systems often apply transmitter-side pre-emphasis (digital pre-equalization) to shape the waveform before the DAC, thereby flattening the effective end-to-end response¹⁶¹. This approach is critical when the receiver must remain simple or when the link budget is tight and residual ISI directly translates into penalties in the vertical eye openings of multilevel PAM. A closely related family of techniques uses structured precoding to reduce post-cursor ISI at the decision point; a typical example in PAM-based links is Tomlinson–Harashima precoding (THP), which can be viewed as a transmitter-side counterpart of DFE that avoids error propagation at the cost of signal “wrapping” operations and associated implementation constraints¹⁶².

In practice, transmitter-side techniques interact strongly with DAC non-idealities and with any subsequent analog filtering (e.g., after ZOH). This is one reason why classical DSP studies sometimes report optimistic results when the DAC/driver are treated as ideal. In a high-fidelity surrogate, pre-emphasis must be evaluated together with quantization, INL/DNL, and dynamic artifacts (e.g., glitches), because these can partially “undo” the intended spectral shaping or introduce new spectral components that reappear as time-domain distortion¹⁶³.

¹⁶⁰ Xi. Pang, O. Ozolins, R. Lin, L. Zhang, A. Udalcovs, L. Xue, R. Schatz, U. Westergren, S. Xiao, W. Hu, G. Jacobsen, S. Popov, & J. Chen, “200 Gb/s Per Lane IM/DD Technologies for Short Reach Optical Interconnects”, *Journal of Lightwave Technology*, vol. 38, pp. 492–503, 2020.

¹⁶¹ *Ibidem*.

¹⁶² T. Wettlin, S. Calabrò, T. Rahman, J. Wei, and N. Stojanovic, “DSP for High-Speed Short-Reach IM/DD Systems Using PAM”, *Journal of Lightwave Technology*, vol. 38, pp. 6771–6778, 2020.

¹⁶³ Xi. Pang, O. Ozolins, R. Lin, L. Zhang, A. Udalcovs, L. Xue, R. Schatz, U. Westergren, S. Xiao, W. Hu, G. Jacobsen, S. Popov, & J. Chen, “200 Gb/s Per Lane IM/DD Technologies for Short Reach Optical Interconnects”, *Journal of Lightwave Technology*, vol. 38, pp. 492–503, 2020.

3.5.1.3 Sequence estimation (MLSE/Viterbi) for severe ISI.

When linear equalization is insufficient, especially under substantial bandwidth limitation, maximum likelihood sequence estimation (MLSE) can offer significant performance gains by exploiting the fact that the received samples depend on multiple past symbols. In PAM links, MLSE is often implemented via the Viterbi algorithm on a finite-memory channel model. Because full-state MLSE becomes rapidly complex as memory grows, reduced-state MLSE variants are commonly used to balance performance and feasibility in short-reach IM/DD receivers¹⁶⁴. These approaches are particularly relevant in scenarios where aggressive spectral shaping or partial-response signaling is used, because the receiver must then resolve intentionally correlated symbol sequences without incurring excessive error floors¹⁶⁵.

3.5.1.4 Weakly nonlinear equalization: Volterra- and polynomial-type methods.

Beyond pure linear ISI, practical IM/DD links exhibit nonlinearities from mixed-signal and electro-optic components (e.g., DAC/driver transfer curves, modulator nonlinearity, and, in amplified links, SOA gain dynamics). A widely used “classical” compromise between transparency and expressive power is to extend linear equalizers into weakly nonlinear models. In the literature on short-reach PAM4, this includes Volterra-series equalizers and related memory-polynomial structures that introduce controlled nonlinear terms (e.g., quadratic/cubic combinations of delayed samples) while retaining an interpretable

¹⁶⁴ Y. Yu, Y. Che, T. Bo, D. Kim, and H. Kim, “Reduced-state MLSE for an IM/DD system using PAM modulation”, *Optics Express*, vol. 28, pp. 38505–38515, 2020.

¹⁶⁵ J. Li, S. H. An, Q. Zhu, X. Li, and Y. Su, “VSB Modified Duobinary PAM4 Signal Transmission in an IM/DD System With Mitigated Image Interference”, *IEEE Photonics Technology Letters*, vol. 32, pp. 363–366, 2020.

filter-like structure¹⁶⁶. Such equalizers often reduce nonlinear distortion contributions to EVM while keeping complexity below that of fully data-driven models. Significantly, their benefit depends on whether the underlying impairments are well approximated by low-order nonlinear expansions over the relevant memory depth. This assumption can break down when the link exhibits strong device memory (e.g., SOA carrier dynamics) or severe quantization/glitch behavior¹⁶⁷.

3.5.1.5 DSP considerations specific to BPAM-style receivers.

For BPAM-like schemes, the receiver DSP can include additional reconstruction steps beyond conventional PAM slicing, because the effective information-bearing waveform is not simply a nonnegative intensity sequence. In particular, the feasibility of direct detection of bipolar constellations relies on oversampling and a reconstruction mechanism that recovers the sign information of the underlying field from intensity samples¹⁶⁸. This makes the overall DSP chain more sensitive to timing, bandwidth, and front-end distortions that reshape the waveform at sub-symbol resolution. From the viewpoint of this thesis, this reinforces the need for a time-resolved surrogate: classical DSP blocks (FFE/DFE, timing recovery, and possibly sequence estimation) must operate on waveforms shaped by realistic DAC/ADC behavior and any optical/electrical nonlinearities, because the reconstruction quality and decision margins can be materially affected by distortions that are invisible in symbol-rate abstractions¹⁶⁹.

3.5.1.6 Positioning relative to ML-based methods.

Classical DSP provides the baseline algorithmic toolbox for short-reach PAM systems and establishes the performance/complexity trade-offs that ML-based

¹⁶⁶ T. Wettlin, S. Calabrò, T. Rahman, J. Wei, and N. Stojanovic, “*DSP for High-Speed Short-Reach IM/DD Systems Using PAM*”, *Journal of Lightwave Technology*, vol. 38, pp. 6771–6778, 2020.

¹⁶⁷ Xi. Pang, O. Ozolins, R. Lin, L. Zhang, A. Udalcovs, L. Xue, R. Schatz, U. Westergren, S. Xiao, W. Hu, G. Jacobsen, S. Popov, & J. Chen, “*200 Gb/s Per Lane IM/DD Technologies for Short Reach Optical Interconnects*”, *Journal of Lightwave Technology*, vol. 38, pp. 492–503, 2020.

¹⁶⁸ M. Secondini and E. Forestieri, “*Direct Detection of Bipolar Pulse Amplitude Modulation*”, *Journal of Lightwave Technology*, 2020.

¹⁶⁹ *Ibidem* 167. (X. Pang *et al.*)

solutions seek to improve. In many recent PAM4 experiments and technology outlooks, classical DSP remains the primary enabler at scale, while ML methods are introduced selectively (e.g., as nonlinear compensators) when classical models exhibit systematic residual errors¹⁷⁰. For an end-to-end digital surrogate intended for optimization, this classical toolbox is not “competing” with ML; instead, it defines (i) the structured signal-processing interfaces where ML can be inserted (e.g., nonlinear post-equalizer, adaptive pre-distortion), and (ii) the reference performance against which any learned method must be judged in a physically grounded way¹⁷¹.

3.5.2 ML-based equalizers and pre-distortion

Beyond classical linear and polynomial DSP, a large body of recent work investigates *machine learning* (ML) models as *data-driven inverse operators* for bandwidth-limited, nonlinear IM/DD links. The motivation is practical: several dominant impairments in short-reach PAM/bPAM chains exhibit *memory* (bandwidth-induced ISI, electrical filtering, SOA dynamics, PD/TIA/AFE responses) and *nonlinearities* (electro-optic transfer functions, saturation/clipping, quantization effects), while accurate closed-form inverses are rarely available across operating points. In this setting, ML-based processing is attractive because it can learn an input-output correction map directly from labeled training sequences without requiring an explicit parametric model for each block^{172,173}.

¹⁷⁰Xi. Pang, O. Ozolins, R. Lin, L. Zhang, A. Udalcovs, L. Xue, R. Schatz, U. Westergren, S. Xiao, W. Hu, G. Jacobsen, S. Popov, & J. Chen, “200 Gb/s Per Lane IM/DD Technologies for Short Reach Optical Interconnects”, *Journal of Lightwave Technology*, vol. 38, pp. 492–503, 2020.

¹⁷¹T. Wettlin, S. Calabrò, T. Rahman, J. Wei, and N. Stojanovic, “DSP for High-Speed Short-Reach IM/DD Systems Using PAM”, *Journal of Lightwave Technology*, vol. 38, pp. 6771–6778, 2020.

¹⁷²D. Zibar, M. Piels, R. Jones, and C. Schaeffer, “Machine Learning Techniques in Optical Communication,” *Journal of Lightwave Technology*, vol. 34, pp. 1442-1452, 2016.

¹⁷³H. Zhou, Y. Li, Y. Liu, L. Yue, C. Gao, W. Li, J. Qiu, et al., “Recent Advances in Equalization Technologies for Short-Reach Optical Links Based on PAM4 Modulation: A Review,” *Applied Sciences*, 2019.

3.5.2.1 Receiver-side ML equalizers.

Most ML equalizers for IM/DD can be viewed as nonlinear filters that approximate:

$$\hat{s}[k] = \mathcal{F}_\theta(\mathbf{y}_k),$$

where $\mathbf{y}_k$ is a window of received samples (possibly multi-rate or feature-augmented) and $\mathcal{F}_\theta$ is a learned function (e.g., DNN/CNN/RNN) with parameters θ . Compared with classical FFE/DFE or Volterra filters, these models can represent richer nonlinear interactions and longer effective memory with fewer hand-designed basis functions, at the cost of training complexity and reduced interpretability¹⁷⁴.

A common architectural split is between (i) *feed-forward DNN/CNN equalizers*, which learn static nonlinear mappings over finite input windows, and (ii) *recurrent models* (RNN/LSTM/GRU and variants), which are naturally suited to channels with longer memory and time-varying behavior. For example, complex-valued recurrent neural network equalizers have been reported as low-complexity solutions for high-rate PAM-4 IM/DD transmission, explicitly exploiting sequence structure rather than treating samples as independent¹⁷⁵. In parallel, sparse or complexity-constrained DNN designs have been explored to reduce inference cost while retaining nonlinear mitigation capability, reflecting a key deployment constraint in cost-sensitive short-reach receivers¹⁷⁶.

Two points are worth emphasizing for an IM/DD state-of-the-art discussion:

1. **Data representation matters.** ML equalizers can operate on symbol-rate samples, oversampled waveforms, or engineered features (e.g., tap-delayed samples, intensity/phase proxies, or outputs of simple linear prefilters). The chosen representation implicitly encodes assumptions about

¹⁷⁴ D. Zibar, M. Piels, R. Jones, and C. Schäffer, “Machine Learning Techniques in Optical Communication,” *Journal of Lightwave Technology*, vol. 34, pp. 1442-1452, 2016.

¹⁷⁵ W. Sheng, C. Liu, J. Xiao, L. Sun, Y. Cai, H. Y. Fu, Q. Li, & G. N. Liu, “Complex-Valued Recurrent Neural Network Equalizer with Low Complexity for a 120-Gbps 50-km Optical PAM-4 IM/DD System,” *Optics Express*, vol. 32, no. 12, pp. 18763-18774, 2024.

¹⁷⁶ G. S. Yadav, C. Chuang, K. Feng, J. Chen, & Y. Chen, “Complexity-Efficient Sparse DNN Nonlinear Equalization for IM/DD 112 Gbps PAM4 Inter-Data Center Optical Interconnects,” *Optics Express*, vol. 32, no. 1, pp. 72-84, 2024.

where nonlinearity and memory “live” in the chain (e.g., predominantly electrical front-end vs. mixed-signal distortions). This is one reason ML literature sometimes reports substantial gains even with limited physical modeling: the model can absorb front-end mismatch and residual ISI, provided that the training data span the relevant regimes^{177,178,179}.

2. Training and generalization are the real bottlenecks. Laboratory links drift (due to temperature, bias points, aging, and connector changes), and IM/DD operation often spans multiple rates and launch powers. Consequently, a recurring theme is to reduce the cost of re-training and improve generalization across conditions. Recent work explores few-shot/small-data adaptation strategies, in which an ML equalizer is adapted to new channel states using limited pilot data, thereby reducing the operational burden relative to full retraining¹⁸⁰.

3.5.2.2 Transmitter-side ML pre-distortion and end-to-end learning.

ML is also increasingly used on the transmitter side, either as explicit *digital pre-distortion* (DPD) or as part of *end-to-end learned transceivers*. In the DPD view, the transmitter applies a learned inverse mapping $x_{\text{pd}} = \mathcal{G}_{\phi}(x)$ so that the *cascade* of DAC + driver + modulator + channel produces a waveform closer to the desired target at the receiver. This is conceptually analogous to radio Frequency power amplifier (RF PA) predistortion. Still, the relevant nonlinearities and memory effects in optical IM/DD differ (e.g., electro-optic transfer, bandwidth

¹⁷⁷ D. Zibar, M. Piels, R. Jones, and C. Schaeffer, “Machine Learning Techniques in Optical Communication,” *Journal of Lightwave Technology*, vol. 34, pp. 1442-1452, 2016.

¹⁷⁸ W. Sheng, C. Liu, J. Xiao, L. Sun, Y. Cai, H. Y. Fu, Q. Li, & G. N. Liu, “Complex-Valued Recurrent Neural Network Equalizer with Low Complexity for a 120-Gbps 50-km Optical PAM-4 IM/DD System,” *Optics Express*, vol. 32, no. 12, pp. 18763-18774, 2024.

¹⁷⁹ G. S. Yadav, C. Chuang, K. Feng, J. Chen, & Y. Chen, “Complexity-Efficient Sparse DNN Nonlinear Equalization for IM/DD 112 Gbps PAM4 Inter-Data Center Optical Interconnects,” *Optics Express*, vol. 32, no. 1, pp. 72-84, 2024.

¹⁸⁰ S. Lennard, F. Barbosa, & F. Ferreira, “Zero-Shot ML Equalization in Coherent Optical Transmission: A First Experimental Demonstration”. *Journal of Lightwave Technology*, 43, 6194-6202, 2025.

limits, quantization, and potentially SOA dynamics). While classical predistortion often relies on structured models (e.g., memory polynomials, Volterra series, LUTs), ML-based predistortion replaces the assumed structure with a learned mapping that can incorporate nonlinear memory in a compact form when training data are available^{181,182}.

A more “integrated” approach is *end-to-end learning*, in which the transmitter and receiver are optimized jointly using differentiable (or differentiable-surrogate) link models. In this family, the system is trained to minimize a task loss (e.g., cross-entropy over symbols, or an EVM/BER proxy), and the learned transmitter can be interpreted as a form of *task-aware predistortion and shaping* that cooperates with the learned receiver¹⁸³. Variants tailored to IM/DD interconnects have been proposed using autoencoder formulations that incorporate channel knowledge or coding structure, aiming to improve robustness while retaining the benefits of joint optimization¹⁸⁴. More broadly, autoencoder-optimized PAM IM/DD transceivers have been demonstrated in amplified IM/DD settings, illustrating how ML training can co-design signaling and receiver processing under non-ideal front-end conditions¹⁸⁵.

3.5.2.3 Where ML equalization overlaps with predistortion.

¹⁸¹ D. Zibar, M. Piels, R. Jones, and C. Schäeffler, “*Machine Learning Techniques in Optical Communication*,” Journal of Lightwave Technology, vol. 34, pp. 1442-1452, 2016.

¹⁸² H. Zhou, Y. Li, Y. Liu, L. Yue, C. Gao, W. Li, J. Qiu, et al., “*Recent Advances in Equalization Technologies for Short-Reach Optical Links Based on PAM4 Modulation: A Review*,” Applied Sciences, 2019.

¹⁸³ B. Karanov, M. Chagnon, F. Thouin, et al., “*End-to-End Deep Learning of Optical Communication*,” Journal of Lightwave Technology, vol. 36, pp. 4843-4851, 2018.

¹⁸⁴ Y. Mu, H. He, J. Ming, J. Song, Q. Xu, R. Gao, Z. Li, et al., “*End-to-End Learning Strategy with Channel-Aided Polar Autoencoder in IM/DD Optical Interconnection*,” Optics Express, vol. 32, no. 22, pp. 39727-39733,

¹⁸⁵ A. Omid, M. Banawan, E. Weckenmann, B. Paquin, A. Geravand, Z. Zheng, “*Auto-Encoder Optimized PAM IM/DD Transceivers for Amplified Fiber Links*,” Journal of Lightwave Technology, vol. 43, pp. 547-556, 2024.

In practice, the boundary between ML equalization and ML predistortion can blur:

- If $\mathcal{F}_\theta$ (receiver) is trained alone, it acts as a *post-compensator* and absorbs residual front-end distortions.
- If $\mathcal{G}_\phi$ (transmitter) is trained alone, it acts as *predistortion/pre-emphasis*, shaping the launched waveform to counteract downstream impairments.
- If both are trained jointly (autoencoder-style), the pair implements an *end-to-end learned inverse split across Tx/Rx*, and “predistortion” becomes one part of a co-optimized transceiver strategy¹⁸⁶.

This overlap is particularly relevant in short-reach IM/DD because some “receiver limitations” (e.g., ADC quantization range or front-end bandwidth) can be partially mitigated by transmitter-side shaping, but only if the model captures the mixed-signal realities governing waveform formation.

Relevance to high-fidelity surrogates. Finally, it is important to connect ML practice to modeling fidelity. Many ML equalizer demonstrations implicitly rely on the fact that *training data already contain* hardware effects (DAC/ADC nonlinearities, ZOH-induced spectral images, front-end filtering, clipping), even when the simulation or analytical model used in the study is simplified. However, this also exposes a limitation: when training data coverage is narrow, ML models can become brittle under changes in time resolution, sampling phase, bias/drift, or component replacement. A high-fidelity, modular surrogate that makes the post-DAC and post-ADC waveform formation explicit (including ZOH and distortion mechanisms) can therefore serve two complementary roles: (i) as a *controlled generator* of diverse, labeled datasets for ML training and stress-testing, and (ii) as

¹⁸⁶ Y. Mu, H. He, J. Ming, J. Song, Q. Xu, R. Gao, Z. Li, et al., “End-to-End Learning Strategy with Channel-Aided Polar Autoencoder in IM/DD Optical Interconnection,” *Optics Express*, vol. 32, no. 22, pp. 39727-39733, 2024.

a *diagnostic scaffold* that helps interpret when an ML equalizer is compensating physical impairments versus exploiting dataset-specific artifacts^{187,188,189}.

3.5.3 Current practice: “black-box” ML not tightly tied to a rich system surrogate

A recurring pattern in the recent literature is that machine-learning (ML) equalizers, classifiers, and end-to-end transceiver optimizers are often trained and evaluated in a *black-box* manner: the optical link is treated as an unknown input-output mapping, and the learning task is framed as “*fit a function that minimizes an error metric.*”¹⁹⁰ This paradigm has produced strong demonstrations in both coherent and IM/DD settings (including NN-based equalization and autoencoder-style end-to-end optimization). Still, it also introduces limitations that become critical when the goal is *engineering-grade, end-to-end optimization* under realistic device constraints¹⁹¹.

Black-box ML typically optimizes *performance* without explicitly representing *causes*. In many studies, the learned model implicitly absorbs a mixture of impairments, such as bandwidth-induced ISI, nonlinear transfer functions, quantization artifacts, SOA dynamics (when present), and noise, without an explicit separation of mechanisms.¹⁹² This is often acceptable when the research objective is to demonstrate feasibility (e.g., “NN equalization improves BER/EVM”), but it

¹⁸⁷ D. Zibar, M. Piels, R. Jones, and C. Schaeffer, “*Machine Learning Techniques in Optical Communication*,” Journal of Lightwave Technology, vol. 34, pp. 1442–1452, 2016.

¹⁸⁸ H. Zhou, Y. Li, Y. Liu, L. Yue, C. Gao, W. Li, J. Qiu, et al., “*Recent Advances in Equalization Technologies for Short-Reach Optical Links Based on PAM4 Modulation: A Review*,” Applied Sciences, 2019.

¹⁸⁹ S. Lennard, F. Barbosa, & F. Ferreira, “*Zero-Shot ML Equalization in Coherent Optical Transmission: A First Experimental Demonstration*”. Journal of Lightwave Technology, 43, 6194–6202, 2025.

¹⁹⁰ D. Zibar, M. Piels, R. Jones, and C. G. Schaeffer, “*Machine Learning Techniques in Optical Communication Systems*”, Journal of Lightwave Technology, vol. 34, no. 6, pp. 1442–1452, 2016.

¹⁹¹ B. Karanov, M. Chagnon, F. Thouin, et al., “*End-to-End Deep Learning of Optical Communication*,” Journal of Lightwave Technology, vol. 36, pp. 4843–4851, 2018.

¹⁹² *Ibidem* 190

becomes a weakness when the objective shifts toward *predictive system design*, parameter sensitivity analysis, and reproducible optimization across hardware revisions and operating points¹⁹³.

A second, closely related issue is *domain shift* between training conditions and deployment conditions. Many ML pipelines rely on either (i) simulated waveforms from simplified link models, or (ii) experimentally collected datasets at a limited set of operating points¹⁹⁴. In the first case, the network may learn the quirks of the simulator rather than the physics of the link; in the second case, the network may overfit to a specific lab configuration (device bias, temperature, optical launch power, ADC settings, SOA operating point, etc.)¹⁹⁵. Both situations can lead to fragile generalization: performance degrades when the channel changes (aging, drift, component replacement), when the baud rate or shaping strategy changes, or when the operating point shifts (e.g., due to power budget variations, SOA gain compression regime)¹⁹⁶.

A third limitation is *data and coverage requirements*. In optical links, the impairment space is inherently multi-dimensional: transmitter DSP settings, DAC/ADC sampling and non-idealities, electrical bandwidth limits, modulator transfer function, fiber/channel conditions, amplifier dynamics, and receiver DSP settings can all interact¹⁹⁷. Purely data-driven training becomes expensive because covering this space requires either extensive lab sweeps (costly and time-consuming) or synthetic datasets (which reintroduce simulation bias). This is particularly relevant for *end-to-end optimization* approaches such as autoencoders:

¹⁹³ A. Omid, M. Banawan, E. Weckenmann, B. Paquin, A. Geravand, Z. Zheng, “*Auto-Encoder Optimized PAM4/DD Transceivers for Amplified Fiber Links*,” *Journal of Lightwave Technology*, vol. 43, pp. 547–556, 2024.

¹⁹⁴ B. Karanov, M. Chagnon, F. Thouin, et al., “*End-to-End Deep Learning of Optical Communication*,” *Journal of Lightwave Technology*, vol. 36, pp. 4843–4851, 2018.

¹⁹⁵ *Ibidem* 193 (A.Omid et al.)

¹⁹⁶ W. S. Saif, M. A. Esmail, A. M. Ragheb, T. A. Alshaw, S. A. Alshebeili, “*Machine Learning Techniques for Optical Performance Monitoring and Modulation Format Identification: A Survey*”, *IEEE Communications Surveys & Tutorials*, Vol. 22, Issue: 4, pp. 2839 – 2882, 2020.

¹⁹⁷ F. N. Khan, Q. Fan, C. Lu, & A. P. T. Lau, “*Machine Learning Methods for Optical Communication Systems and Networks*”, *Optical Fiber Telecommunications VII*, pp. 921-978, 2020.

their success depends strongly on how faithfully the training environment reflects the deployment environment, and on whether the optimization targets remain stable under realistic component variability¹⁹⁸.

A fourth issue is *interpretability and controllability*, which directly impacts thesis-level claims about *why* a method works and *where* it is applicable. Black-box models can be rigorously benchmarked using BER/EVM. Still, they do not automatically provide engineering insight, such as: which block dominates penalty, which parameter is most sensitive, what trade-off is being made between linear and nonlinear compensation, or how a learned equalizer behaves when the receiver front-end saturates, or the SOA enters deeper compression¹⁹⁹. This matters even more in IM/DD research where “simplified” receivers are a key design goal. If the receiver relies on oversampling and reconstruction mechanisms (as in BPAM-like approaches), the learned model may exploit subtle waveform features that are not robust under realistic analog distortions unless those distortions are represented during training²⁰⁰.

Recent work has therefore begun to move away from purely black-box learning toward *physics- or optics-informed ML*, in which known structure is embedded in the model class, the training pipeline, or the forward model²⁰¹. In IM/DD specifically, optics-informed neural-network approaches have been proposed to retain high end-to-end performance while improving robustness by incorporating link structure into the learning process, thereby reducing the burden on purely data-driven fitting²⁰². In parallel, the optical-networks community has emphasized that

¹⁹⁸ H. Klopries, D. O. S. Torres, A. Schwung, “*Synthetic time series dataset generation for unsupervised autoencoders*,” 2022 IEEE 27th International Conference on Emerging Technologies and Factory Automation (ETFA), 1-8. 2022.

¹⁹⁹ D. Zibar, M. Piels, R. Jones, and C. G. Schäeffler, “*Machine Learning Techniques in Optical Communication Systems*”, Journal of Lightwave Technology, vol. 34, no. 6, pp. 1442–1452, 2016.

²⁰⁰ B. Karanov, M. Chagnon, F. Thouin, et al., “*End-to-End Deep Learning of Optical Communication*,” Journal of Lightwave Technology, vol. 36, pp. 4843–4851, 2018.

²⁰¹ I. Roumpos, L. De Marinis, M. Kirtas, N. Passalis, A. Tefas, G. Contestabile, N. Pleros, M. Moralis-Pegios, & K. Vyrsokinos, “*High-Performance End-to-End Deep Learning IM/DD Link Using Optics-Informed Neural Networks*”, Optics Express, vol. 31, no. 12, pp. 20068–20079, 2023.

²⁰² *Ibidem*.

ML must be integrated with calibrated models and monitoring/telemetry to be operationally meaningful, rather than remaining a standalone predictor trained on narrow datasets²⁰³.

From the viewpoint of this thesis, these observations motivate a clear methodological position: *ML should not be asked to compensate for what the model refuses to represent*. Instead, a *rich digital surrogate*, with explicit device-level non-idealities, correct sampling/time resolution, and calibrated operating points, can be used to (i) generate diverse and physically meaningful training data, (ii) expose which impairments dominate under which regimes, and (iii) support ML optimization that remains consistent when moved from simulation to experiment. This linkage does not remove the need for ML; rather, it makes ML *more reliable* by anchoring learning to realistic end-to-end behavior and by enabling systematic stress-testing across operating conditions that would otherwise be impractical to cover experimentally²⁰⁴.

3.6 Gap analysis and position of this thesis

3.6.1 Gap analysis: why an end-to-end, device-aware surrogate is still missing

The preceding sections highlight a recurring structural pattern in the IM-DD literature: high-level link studies and standards-facing evaluations typically work with symbol-rate abstractions and a small set of “dominant” impairments (e.g., bandwidth-limited ISI and receiver noise), while component-focused studies provide detailed device physics but are rarely integrated into an end-to-end chain suitable for system optimization and learning. This split is understandable because each layer answers a different engineering question. Still, it becomes a bottleneck when the research objective is the end-to-end optimization of practical PAM/BPAM links under realistic mixed-signal constraints, where DAC/ADC

²⁰³ W. S. Saif, M. A. Esmail, A. M. Ragheb, T. A. Alshaw, S. A. Alshebeili, “*Machine Learning Techniques for Optical Performance Monitoring and Modulation Format Identification: A Survey*”, IEEE Communications Surveys & Tutorials, Vol. 22, Issue: 4, pp. 2839 – 2882, 2020.

²⁰⁴ A. Omid, M. Banawan, E. Weckenmann, B. Paquin, A. Geravand, Z. Zheng, “*Auto-Encoder Optimized PAM IM/DD Transceivers for Amplified Fiber Links*,” Journal of Lightwave Technology, vol. 43, pp. 547–556, 2024.

artifacts, amplifier dynamics, and receiver-front-end bandwidth and noise interact strongly^{205,206}.

A helpful way to summarize the current landscape is a fidelity “ladder”: (i) fast analytical or semi-analytical link models for quick trends, (ii) waveform-level block simulations that capture DSP and filtering but often idealize mixed-signal hardware, (iii) device-/circuit-aware co-simulation calibrated by measured responses, and (iv) operational digital twins that are continuously synchronized with a deployed system for monitoring, prediction, and control^{207,208}. The key observation for this thesis is that most IM-DD PAM research occupies levels (i)-(ii), while digital twin work in optical communications often targets levels (iii)-(iv) but at network/operations scope (fault management, configuration, transmission prognosis) rather than transmitter/receiver hardware non-idealities at sample-level resolution^{209,210}.

²⁰⁵ X. Zhou, R. Urata, and H. Liu, “Beyond 1 Tb/s Intra-Data Center Interconnect Technology: IM-DD OR Coherent?”, *Journal of Lightwave Technology*, vol. 38, no. 2, pp. 475-484, 2020.

²⁰⁶ T. Wettlin, S. Calabrò, T. Rahman, J. Wei, and N. Stojanovic, “DSP for High-Speed Short-Reach IM/DD Systems Using PAM”, *Journal of Lightwave Technology*, vol. 38, no. 24, pp. 6771-6778, 2020.

²⁰⁷ W. Kritzinger, M. Karner, G. Traar, J. Henjes, and W. Sihn, “Digital Twin in manufacturing: A categorical literature review and classification”, *IFAC-PapersOnLine*, vol. 51, no. 11, pp. 1016-1022, 2018.

²⁰⁸ B. R. Barricelli, E. Casiraghi, and D. Fogli, “A survey on digital twin: definitions, characteristics, applications, and design implications”, *IEEE Access*, vol. 7, pp. 167653-167671, 2019.

²⁰⁹ R. Vilalta, L. Gifre, R. Casellas, R. Martínez, R. Muñoz, A. Mozo, “Applying Digital Twins to Optical Networks with Cloud-Native SDN Controllers”, *IEEE Communications Magazine*, pp. 129-133, 2023

²¹⁰ D. Wang, Z. Zhang, M. Zhang, M. Fu, J. Li, S. Cai, “The Role of Digital Twin in Optical Communication: Fault Management, Hardware Configuration, and Transmission Simulation”, *IEEE Communications Magazine*, pp. 133-139, 2021.

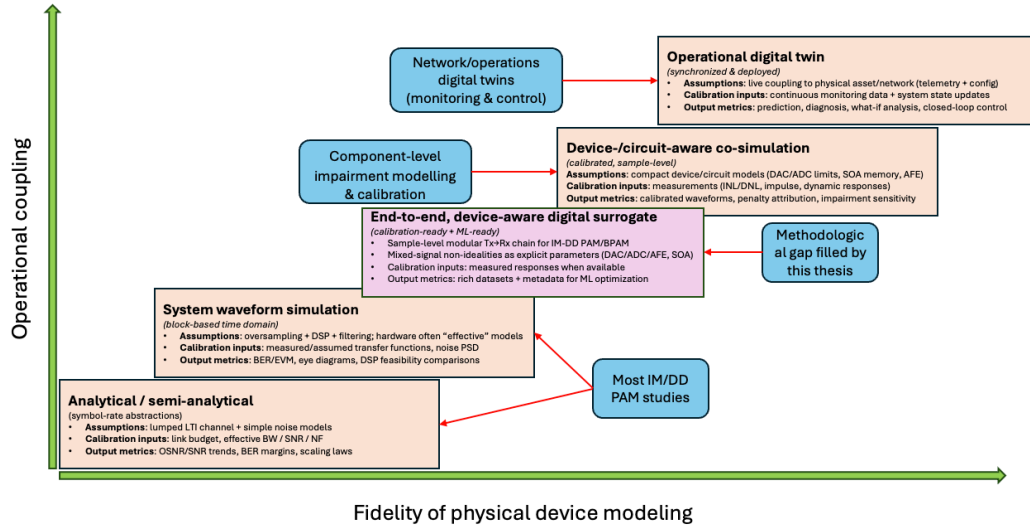


Figure 3.1 Fidelity Ladder for modelling IM-DD links

From the IM-DD system side, many “system-level” PAM studies intentionally collapse complex hardware into a small number of linear, time-invariant transfer functions and lumped noise sources, because these abstractions enable wide parameter sweeps and a clean comparison of equalization strategies at scale²¹¹. Even when emerging formats such as bipolar PAM are considered, the emphasis is commonly placed on the feasibility of recovery and the DSP principle (e.g., oversampling and reconstruction), rather than on a calibrated device-level surrogate that explains how the feasibility window changes with DAC static/dynamic errors, SOA memory, or front-end sampling/quantization constraints²¹². In parallel, amplifier literature provides rich descriptions of SOA gain dynamics, amplitude-phase coupling, and intraband effects that matter for multilevel IM-DD waveforms (and can create memory and nonlinear distortion). Still, those models are often used in isolation or under assumptions that are not aligned with a full mixed-signal Tx→Rx surrogate intended for ML data generation²¹³.

²¹¹ T. Wettlin, S. Calabrò, T. Rahman, J. Wei, and N. Stojanovic, “*DSP for High-Speed Short-Reach IM/DD Systems Using PAM*”, *Journal of Lightwave Technology*, vol. 38, no. 24, pp. 6771–6778, 2020.

²¹² M. Secondini and E. Forestieri, “*Direct Detection of Bipolar Pulse Amplitude Modulation*”, *Journal of Lightwave Technology*, 2020.

²¹³ A. Sobhanan, A. Anthur, S. O’Duill, M. Pelusi, S. Namiki, L. Barry, D. Venkitesh, and G. P. Agrawal, “*Semiconductor optical amplifiers: recent advances and applications*”, *Advances in Optics and Photonics*, vol. 14, no. 3, pp. 571–651, 2022.

From the “digital twin/surrogate” side, the core definitions are converging: a digital twin is typically described as a digital representation of a physical asset/process with explicit coupling to the physical system (often bidirectional and time-evolving), whereas a digital surrogate is frequently used as a predictive stand-in model that may be calibrated but does not necessarily maintain continuous synchronization^{214,215}. In optical communications, many works labeled “digital twin” focus on operational objectives such as network fault management, configuration, and transmission-simulation-as-a-service, in which the twin ingests telemetry and predicts network-level behavior²¹⁶. Other twin-like efforts build ML-driven models for specific monitoring functions (e.g., reflectometry and fault localization), again emphasizing operational accuracy for a particular task rather than modular end-to-end emulation of the complete IM-DD transceiver chain²¹⁷. These contributions are valuable, but they typically do not provide (and do not claim to provide) a reusable, block-modular transmitter/receiver surrogate that exposes device non-idealities as controllable parameters for optimization studies.

This is the concrete gap that motivates the thesis. There is no widely adopted, end-to-end, modular digital surrogate for IM-DD PAM/BPAM links that simultaneously (i) represents realistic mixed-signal behavior of the DAC/driver/modulator and ADC/AFE, (ii) includes amplifier non-idealities (e.g., SOA gain saturation and memory) in a calibration-ready form, and (iii) is explicitly designed to generate rich, structured training data for ML-based equalization and pre-distortion across operating points. While isolated pieces exist, high-quality DSP benchmarking at the system level, device-accurate component models, and operational digital-twin frameworks, these strands remain largely disconnected in

²¹⁴ W. Kritzinger, M. Karner, G. Traar, J. Henjes, and W. Sihn, “*Digital Twin in manufacturing: A categorical literature review and classification*”, IFAC-PapersOnLine, vol. 51, no. 11, pp. 1016-1022, 2018.

²¹⁵ B. R. Barricelli, E. Casiraghi, and D. Fogli, “*A survey on digital twin: definitions, characteristics, applications, and design implications*”, IEEE Access, vol. 7, pp. 167653-167671, 2019.

²¹⁶ D. Wang, Z. Zhang, M. Zhang, M. Fu, J. Li, S. Cai, “*The Role of Digital Twin in Optical Communication: Fault Management, Hardware Configuration, and Transmission Simulation*”, IEEE Communications Magazine, pp. 133-139, 2021.

²¹⁷ D. Sequeira, M. Ruiz, N. Costa, A. Napoli, J. Pedro, & L. Velasco, “*OCATA: a deep-learning-based digital twin for the optical time domain analyzer*”, Journal of Optical Communications and Networking, vol. 15, no. 2, pp. 87-97, 2023.

the literature, leaving a clear opportunity for a unified surrogate that is both physically meaningful and ML-ready^{218,219,220}.

Literature strand	Typical scope	Typical modeling level	Missing for ML-ready surrogate
IM/DD PAM system/DSP studies	End-to-end link	Waveform (effective)	Device realism
Converter-aware studies (DAC/ADC)	Block/component	Sample-level	End-to-end
SOA physics & dynamics	Component	Physics-aware	System integration
Receiver front-end (PD/TIA/AFE)	Component	Circuit/TF	Joint optimization
Bipolar PAM feasibility	Link/DSP concept	Waveform	Hardware constraints
ML equalization / pre-distortion	Task-driven	Data-driven	Physical grounding
Digital twin (optical networks)	Operations	Telemetry-driven	Transceiver detail
This thesis	End-to-end link	Sample-level modular	—

Table 3.1 Literature coverage matrix

3.6.2 How our work fills this gap (very short)

Building on the gaps identified above, this thesis develops an end-to-end *digital surrogate* for short-reach IM-DD PAM/BPAM links that is *device-aware*, *calibration-ready*, and explicitly designed for *optimization and learning*. The surrogate is organized as a block-modular Tx→Rx chain at waveform (sample-level) resolution, so that transmitter and receiver non-idealities are not absorbed into a single “effective channel,” but remain visible as controllable parameters. In particular, it captures realistic mixed-signal behavior in the *DAC/driver/modulator* and *ADC/AFE* (e.g., quantization and sampling constraints, static/dynamic conversion artifacts). It integrates *SOA dynamics and memory* in a form that can be

²¹⁸ T. Wettlin, S. Calabrò, T. Rahman, J. Wei, and N. Stojanovic, “*DSP for High-Speed Short-Reach IM/DD Systems Using PAM*”, Journal of Lightwave Technology, vol. 38, no. 24, pp. 6771-6778, 2020.

²¹⁹ B. R. Barricelli, E. Casiraghi, and D. Fogli, “*A survey on digital twin: definitions, characteristics, applications, and design implications*”, IEEE Access, vol. 7, pp. 167653-167671, 2019.

²²⁰ A. Sobhanan, A. Anthur, S. O’Duill, M. Pelusi, S. Namiki, L. Barry, D. Venkitesh, and G. P. Agrawal, “*Semiconductor optical amplifiers: recent advances and applications*”, Advances in Optics and Photonics, vol. 14, no. 3, pp. 571-651, 2022.

tuned to measured responses, enabling consistent evaluation of multilevel PAM/BPAM under practical hardware limits.

Crucially, the surrogate is not only a simulation tool but also a *data generator*: it produces rich, structured datasets across operating points (symbol rate, OSNR, bandwidth constraints, amplifier settings, and impairment levels) to support ML-based equalization and predistortion, as well as hybrid approaches that combine model-based structure with data-driven adaptation. By unifying system-level DSP evaluation, device-aware impairment modeling, and ML-ready metadata generation within a single coherent framework, the proposed surrogate bridges the gap between isolated component models and abstract link studies and provides a practical foundation for reproducible, end-to-end optimization of IM-DD PAM/BPAM transceivers.

Chapter 4

Architecture of the Digital Surrogate Framework

This chapter formalizes the architecture of the proposed digital surrogate framework as a distinct scientific entity. The main methodological assertion is that, for short-reach IM-DD PAM/BPAM systems, the credibility and usefulness of any end-to-end surrogate depend not only on the accuracy of individual component models but also on the software and data architecture that connects them: modular decomposition, strict parameterization, domain-aware interfaces (digital/electrical/optical), consistent time management, and explicit tracking of impairments and metadata. In many studies, the simulation workflow is built specifically for a single scenario and is not designed for reuse, extension, or systematic calibration; such workflows are often referred to in the literature as *ad hoc* simulations (from Latin *ad hoc*, “for this specific purpose”). By turning these architectural choices into a documented, reproducible framework, rather than an *ad hoc* simulation script, this thesis offers a transferable method for constructing, calibrating, and expanding realistic link surrogates.

The chapter, therefore, focuses on how the framework is organized, how information flows between blocks, and how the architecture enforces truthfulness in implementation: any physical assumption must be explicit, any formula must be traceable to a documented model, and any impairment introduced into the signal chain must have a clear injection point and a controllable parameterization. In this sense, the architecture is not merely a software design; it is the mechanism that renders subsequent modeling results interpretable, comparable across scenarios, and amenable to systematic optimization.

The relationship to Chapters 5-7 is direct: this chapter defines the interfaces and conventions that all subsequent component models must adhere to. Chapters 5 and 6 introduce the transmitter, channel, and receiver modules (including relevant non-idealities) and show how they are integrated into the standard chain without compromising unit consistency, timing consistency, or metadata conventions.

Chapter 7 then builds on the same architectural backbone to support DSP and ML workflows, where training, testing, and transfer across operating points require not only waveforms but also structured context (operating conditions, impairment settings, and calibration state). In other words, Chapter 4 provides the “contract” that makes the subsequent chapters composable. Each block can evolve in fidelity while remaining compatible with the same end-to-end execution, visualization, and validation strategy.

More practically, Chapter 4 also explains how the architecture supports incremental scientific development: blocks can be enabled/disabled, replaced by measured responses, or refined from simplified to device-aware implementations without rewriting the rest of the chain. This is essential for a thesis that must validate ideas progressively: first, ensuring correct signal flow and metrics; then introducing non-idealities; then calibrating to experiments; and finally using the surrogate as a controlled environment for algorithmic and ML studies.

Finally, the framework's target use cases are threefold. First, it supports end-to-end studies of IM-DD PAM/BPAM transceivers at waveform (sample-level) resolution, enabling consistent evaluation of interactions among bandwidth limits, DSP, converter constraints, amplifier memory, and receiver-front-end effects within a single model. Second, it is designed for calibration and hardware alignment: the architecture provides explicit hooks for measured transfer functions and measured device responses, allowing the surrogate to move beyond “idealized blocks” toward a configuration that reflects a specific laboratory setup or component chain. Third, it serves as a machine-learning dataset generator: by running controlled sweeps over operating points and impairment parameters while recording both waveforms and structured metadata, the surrogate produces training and validation data suitable for ML-based equalization, pre-distortion, and hybrid model-driven/data-driven optimization.

Across these use-cases, the key objective is reproducibility: the same framework that generates results also preserves the conditions under which those results were produced. This enables meaningful comparisons between classical DSP and ML approaches, supports sensitivity analysis and ablation studies, and

provides a stable foundation for the component-level modeling and optimization results presented in the remainder of the thesis.

4.1 Design principles

The digital surrogate is designed not as a single-use codebase for a single experiment, but as a reusable methodological framework that can be expanded in fidelity, calibrated to hardware, and used systematically for optimization and learning. The design principles in this section translate the gap identified in Chapter 3 into specific architectural rules: how the surrogate is decomposed, how parameters are handled, how domains are represented, and how outputs and metadata are generated. These rules make the framework flexible: the same architecture that supports short-reach IM-DD PAM(BPAM)¹ can be reused to model other transmission regimes (including coherent links) by replacing or extending blocks while maintaining interfaces, time handling, and metadata standards. Additionally, the design aligns with an emerging trend in optical communications research: treating the communication chain as a structured, configurable pipeline² that supports both traditional DSP evaluation and modern learning-based optimization³, which is only possible when the underlying modeling is modular, well-parameterized, and internally consistent.

4.1.1 Modularity and strict separation into blocks (Tx/channel / Rx)

The first principle is a strict decomposition of the end-to-end chain into well-defined blocks grouped into transmitter (Tx), channel, and receiver (Rx)

¹ X. Pang, O. Ozolins, R. Lin, L. Zhang, A. Udalcovs, L. Xue, R. Schatz, U. Westergren, S. Xiao, W. Hu, G. Jacobsen, S. Popov, and J. Chen, "200 Gbps/Lane IM/DD Technologies for Short Reach Optical Interconnects", *Journal of Lightwave Technology*, vol. 38, no. 2, pp. 492-503, 2020.

² Z. Niu, H. Yang, L. Li, M. Shi, G. Xu, W. Hu & L. Yi, "Learnable digital signal processing benchmark for linearity compensation in optical fiber communications", *Light: Science & Applications*, 2024.

³ Q. Fan, G. Zhou, T. Gui, C. Lu, and A. P. T. Lau, "Advancing theoretical understanding and practical performance of signal processing for nonlinear optical communications through machine learning", *Nature Communications*, vol. 11, article 3694, 2020.

subsystems. This choice is driven by a practical research need: IM-DD performance is dominated by interactions between impairments originating in different physical parts of the chain (e.g., DAC behavior shaping modulator drive, bandwidth limits interacting with equalization, amplifier memory interacting with symbol history). If these effects are embedded in a monolithic model, it becomes difficult to isolate causes, perform controlled ablations, or reuse components across scenarios. A block-modular structure allows each impairment to have a clear injection point and parameterization, enabling systematic “what changes if we modify only this block?” experiments⁴, an essential requirement for scientific interpretability and credible comparisons between DSP/ML strategies⁵.

On the Tx side, the modular decomposition explicitly includes the pre-DAC DSP as part of the transmitter pipeline: symbol mapping, pulse shaping, optional pre-emphasis, and optional predistortion. These elements are not merely implementation details; they determine waveform statistics and spectral occupancy that condition DAC behavior and downstream nonlinearities. Keeping them as explicit blocks enables fair comparisons between classical DSP configurations and learned predistortion strategies⁶ and supports end-to-end optimization approaches where some Tx degrees of freedom (e.g., shaping or bias parameters) can be tuned jointly with receiver processing⁷.

⁴ T. Wettlin, S. Calabrò, T. Rahman, J. Wei, N. Stojanovic, & S. Pachnicke, “*DSP for High-Speed Short-Reach IM/DD Systems Using PAM*”, *Journal of Lightwave Technology*, vol. 38, no. 24, pp. 6771-6778, 2020.

⁵ Z. Xu, T. Ji, Q. Wu, W. Lu, H. Ji, Y. Yang, G. Qiao, J. Tang, C. Cheng, L. Liu, S. Wang, J. Liang, Z. Sun, L. Fan, J. Wei, and W. Shieh, “*Advanced Neural Network-Based Equalization in Intensity-Modulated Direct-Detection Optical Systems: Current Status and Future Trends*”, *Photonics*, vol. 11, no. 8, article 702, 2024.

⁶ Z. Niu, H. Yang, L. Li, M. Shi, G. Xu, W. Hu & L. Yi, “*Learnable digital signal processing benchmark for linearity compensation in optical fiber communications*”, *Light: Science & Applications*, 2024.

⁷ Q. Fan, G. Zhou, T. Gui, C. Lu, and A. P. T. Lau, “*Advancing theoretical understanding and practical performance of signal processing for nonlinear optical communications through machine learning*”, *Nature Communications*, vol. 11, article 3694, 2020.

This modularity also supports expandability and long-term maintenance. As models evolve (e.g., by refining DAC static nonlinearities or replacing a simplified channel response with a richer propagation model), each refinement can be introduced locally within a single block without disrupting the entire chain. This block-oriented organization is also widely adopted in commercial simulation environments, where complex optical links are constructed by interconnecting reusable functional modules. Commercial simulators can be useful for many link-design tasks^{8,9}, but they are not necessarily designed to expose the same level of internal “hooks” (per-block impairment control, systematic metadata extraction, reproducible dataset pipelines) that are necessary for the methodological goals of this thesis. For this reason, the current framework treats each block as a transparent scientific artifact, designed for inspection, modification, and controlled experimentation.

4.1.2 Full parameterization and “no hard-coding” policy

The second principle is full parameterization: every model behavior must be managed through a centralized configuration structure, and no critical parameter should be hard-coded inside blocks. This is essential for two reasons. First, it guarantees reproducibility: every figure or dataset can be linked to a specific parameter set, and repeating the experiment involves executing the same configuration rather than reinterpreting embedded constants. Second, it allows systematic exploration: sensitivity analysis, parameter sweeps, and calibration routines depend on parameters being externally visible and machine-editable. This policy is significant for mixed-signal modeling, where many parameters are “small” but decisive (converter resolution, clipping margins, bandwidth corner frequencies, noise bandwidths, amplifier recovery times). Hard-coding such values makes the

⁸ N. R. Hugar, P. Piruthiviraj, Maleeha, and V. Dhinesh, “*Simulation and Analysis of Bit Error Rate in Optical Fiber Communication Using Optisystem*”, 2024 7th International Conference on Devices, Circuits and Systems (ICDCS), pp. 67-71, 2024.

⁹ A. Atieh and M. Raytchev, “*Optical communication system (OptiSystem) software enabling remote education and teaching*”, Proceedings of SPIE, Seventeenth Conference on Education and Training in Optics and Photonics (ETOP 2023), 2023.

model fragile: results may appear correct while silently reflecting hidden assumptions. In contrast, central parameterization enables the surrogate to support multiple operating regimes (e.g., baud rate, modulation order, noise conditions) and to be extended to other transmission systems by adding new parameter branches (e.g., coherent receiver parameters) without altering the underlying architectural logic. The same policy also enables lightweight execution modes: blocks can be enabled/disabled and metadata gathering toggled without altering code paths, which is important for efficient debugging and for generating large ML datasets in a controlled way¹⁰, an increasingly common requirement in modern optical-system research workflows¹¹.

4.1.3 Domain-aware modeling (digital/electrical/optical) and interface discipline

The third principle is domain-aware modeling, enforced through disciplined interfaces. In IM-DD links, a waveform transitions through fundamentally different physical representations: a discrete-time digital sequence, an electrical waveform at the modulator drive, an optical field or intensity after electro-optic conversion, and an electrical current/voltage after photodetection and analog front-end (AFE) conditioning (where AFE denotes the analog circuitry following the photodiode, typically including transimpedance amplification and filtering stages)¹². If domain boundaries are not explicit, it becomes easy to introduce unit inconsistencies or apply an impairment in the wrong domain (e.g., applying optical noise to an electrical signal, or mixing amplitude scaling conventions).

¹⁰ Z. Niu, H. Yang, L. Li, M. Shi, G. Xu, W. Hu & L. Yi, "*Learnable digital signal processing benchmark for linearity compensation in optical fiber communications*", Light: Science & Applications, 2024.

¹¹ H. Yang, Z. Niu, H. Zhao, S. Xiao, W. Hu, L. Yi, "*Fast and Accurate Waveform Modeling of Long-Haul Multi-Channel Optical Fiber Transmission Using a Hybrid Model-Data Driven Scheme*", Journal of Lightwave Technology, vol. 40, no. 14, pp. 4571-4580, 2022.

¹² X. Pang, O. Ozolins, R. Lin, L. Zhang, A. Udalcovs, L. Xue, R. Schatz, U. Westergren, S. Xiao, W. Hu, G. Jacobsen, S. Popov, and J. Chen, "*200 Gbps/Lane IM/DD Technologies for Short Reach Optical Interconnects*", Journal of Lightwave Technology, vol. 38, no. 2, pp. 492-503, 2020.

To prevent this, each block adheres to a clear interface contract: it declares the input domain, produces an output with a declared output domain, and applies only those impairments meaningful in that domain¹³. This improves scientific truthfulness: when a result changes, it is traceable to a specific physical transformation rather than an ambiguous operation. It also makes the architecture expandable beyond IM-DD: the same discipline extends naturally if optical signals are represented as complex fields and coherent receivers introduce additional domain transformations (e.g., IQ mixing, carrier recovery). The point is not that IM-DD and coherent share the same physics, but that they benefit from the same architectural rule: domain transitions must be explicit and systematically enforced.

4.1.4 Calibration readiness and hardware-aligned abstraction level

The fourth principle is calibration readiness: the surrogate must be able to incorporate measured responses and align its internal parameters with real hardware behavior. This principle is motivated by a central weakness of many purely “clean” simulations: they can be excellent for conceptual trends but fail when used to explain or predict the behavior of a specific laboratory chain, especially when mixed-signal artifacts dominate performance. Calibration readiness, therefore, influences the choice of abstraction level. The surrogate favors compact, hardware-aligned models: detailed enough to reproduce dominant non-idealities and memory effects, but structured so parameters can be identified from measurements (transfer functions, gain curves, response dynamics, noise figures, and converter characteristics).

This principle is also consistent with a broader trend toward hybrid modeling strategies in optical communications, in which data-driven components complement physics-based structures to achieve both accuracy and tractability

¹³ T. Wettlin, S. Calabrò, T. Rahman, J. Wei, N. Stojanovic, & S. Pachnicke, "DSP for High-Speed Short-Reach IM/DD Systems Using PAM", *Journal of Lightwave Technology*, vol. 38, no. 24, pp. 6771-6778, 2020.

across operating points¹⁴. In that context, the architectural requirement is not merely “high fidelity,” but identifiability: a model that is organized so that measured data can be mapped to the correct block parameters, and improvements can be introduced incrementally without restructuring the whole chain.

4.1.5 ML-readiness: structured metadata and repeatable dataset generation

The fifth principle is ML-readiness, meaning that the surrogate is designed from the outset to generate datasets usable for learning and optimization, rather than as an afterthought. In many studies, ML is applied as a black-box equalizer trained on waveforms without sufficient contextual information about the operating point, impairment settings, or calibration state. This limits generalization and makes it hard to interpret what a model has learned. Recent literature increasingly emphasizes that ML effectiveness in optical links depends strongly on how the modeling and data pipeline are constructed¹⁵, particularly when the objective is optimization, robustness across conditions, or extracting insight rather than only curve-fitting¹⁶.

In the present framework, dataset generation is treated as part of the architecture: whenever a waveform is produced, the framework can also generate structured metadata describing the conditions under which it was generated (e.g., rates, sampling, domain, impairment parameters, calibration flags, and block lineage). This supports controlled training across operating points, enables fair comparisons between classical DSP and ML methods¹⁷, and supports hybrid

¹⁴ H. Yang, Z. Niu, H. Zhao, S. Xiao, W. Hu, L. Yi, "Fast and Accurate Waveform Modeling of Long-Haul Multi-Channel Optical Fiber Transmission Using a Hybrid Model-Data Driven Scheme", *Journal of Lightwave Technology*, vol. 40, no. 14, pp. 4571-4580, 2022.

¹⁵ Z. Niu, H. Yang, L. Li, M. Shi, G. Xu, W. Hu & L. Yi, "Learnable digital signal processing benchmark for linearity compensation in optical fiber communications", *Light: Science & Applications*, 2024.

¹⁶ Z. Xu, T. Ji, Q. Wu, W. Lu, H. Ji, Y. Yang, G. Qiao, J. Tang, C. Cheng, L. Liu, S. Wang, J. Liang, Z. Sun, L. Fan, J. Wei, and W. Shieh, "Advanced Neural Network-Based Equalization in Intensity-Modulated Direct-Detection Optical Systems: Current Status and Future Trends", *Photonics*, vol. 11, no. 8, article 702, 2024.

¹⁷ *Ibidem*.

strategies in which learned models exploit physical structure (e.g., by learning residual nonlinearities on top of a known baseline)¹⁸. It also improves expandability: as new impairments or blocks are added, they can be incorporated into the metadata schema without disrupting existing pipelines, enabling continuous growth of the surrogate’s “training universe” while preserving reproducibility and interpretability.

4.2 Global configuration layer

The framework is driven by a single global configuration layer that makes the surrogate *programmable as an experiment*: every block behavior, impairment setting, operating point, and metadata policy is externally defined and can be reproduced, swept, or calibrated without rewriting code. This section describes (i) how the configuration is structured and named, (ii) how complete Tx→channel→Rx “profiles” are activated for different research tasks, and (iii) how reproducibility is enforced through explicit scenario bookkeeping.

4.2.1 The parameters structure: hierarchy and naming conventions

The global configuration is structured as a hierarchical “parameters” structure whose top-level organization mirrors the physical chain and research workflow. At the top level, the structure segregates Tx, channel, and Rx, and includes cross-cutting branches for *visualization/virtual instruments*, *measurement/metrics*, and *ML readiness*. This hierarchy is deliberate: it allows the surrogate to be upgraded in fidelity by *replacing or enhancing a single block* (such as DAC, SOA, channel propagation, ADC) while maintaining consistent interfaces and a stable vocabulary.

Within Tx, the hierarchy reflects the development order and the physical signal flow. It begins with the *bit/source layer* (imported bits versus random bits versus standardized PRBS families, plus explicit control of reproducible seeds versus shuffled runs), then includes mapping and coding placeholders (e.g., Gray coding

¹⁸ Q. Fan, G. Zhou, T. Gui, C. Lu, and A. P. T. Lau, “Advancing theoretical understanding and practical performance of signal processing for nonlinear optical communications through machine learning”, Nature Communications, vol. 11, article 3694, 2020.

policy, reserved hooks for FEC as pass-through until enabled), followed by the pre-DAC DSP chain (symbol mapping, pulse shaping, optional pre-emphasis, optional predistortion). The DAC is represented as a parent module that calls internal subfunctions matching device behavior (e.g., sampling-rate matching and pre-emphasis as digital sub-steps, and downstream mixed-signal effects such as quantization, remapping, nonlinearities, filters, and noise, as configured). The TX chain then advances through the *driver* (gain, bandwidth shaping, ISI contributions, and electrical noise options), *laser/CW source* (including phase-drift options under a bandwidth-aware equivalent representation), and *MZM* (electro-optic transformation driven by the electrical waveform and the bias-control block).

Within the *channel*, the hierarchy is organized by impairment groups that can be turned on or off independently: attenuation, chromatic dispersion, and (where applicable) nonlinear effects based on fiber length and operating conditions. Noise injection is explicitly configurable to match either simulation or laboratory setups: the configuration supports either direct AWGN injection or an ASE-like mechanism consistent with attenuator-amplifier OSNR control workflows, depending on the desired experimental emulation.

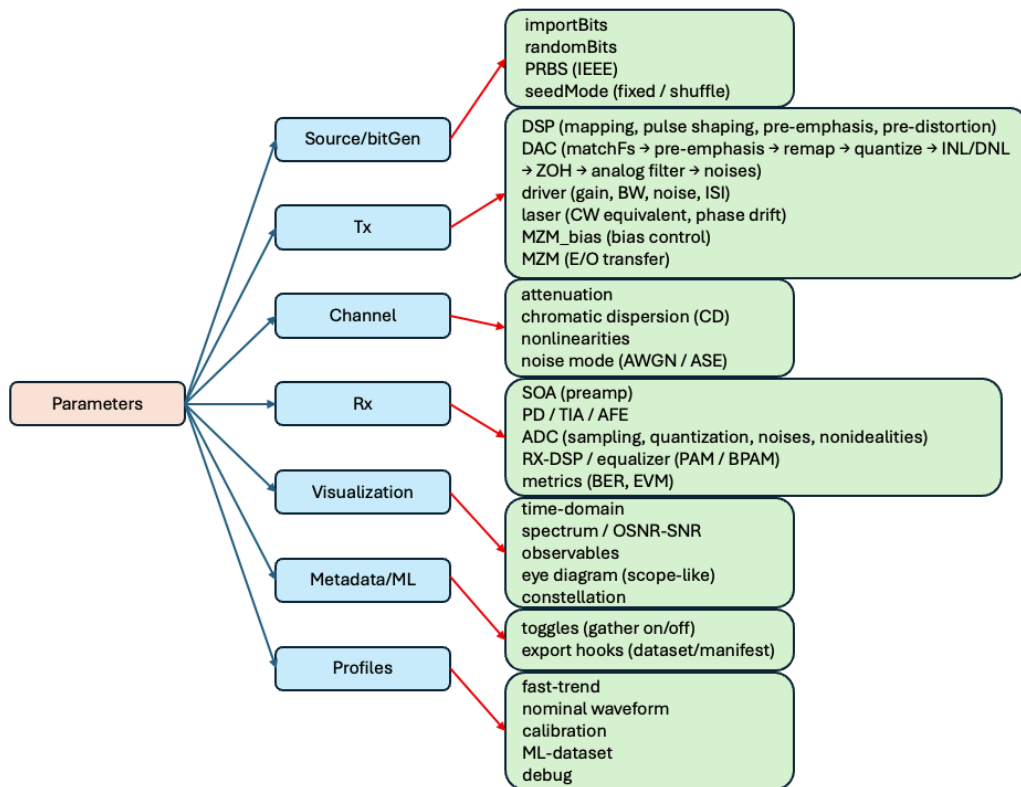


Figure 4.1 Parameters hierarchy map

Within Rx, the hierarchy reflects both the physical front end and the post-processing chain. It usually includes optional *SOA pre-amplification*, the *photodiode or photoreceiver*, the *analog front-end (AFE)* conditioning (i.e., the analog circuitry immediately following photodetection, typically including transimpedance amplification and analog filtering), the *ADC* (as a parent module with internal sub-effects), and finally the *receiver DSP or equalizer* supporting both standard PAM and BPAM processing paths. A dedicated branch defines *measurement functions* (BER/EVM and other waveform-quality metrics), so that evaluation is treated as a first-class, configurable part of the workflow rather than ad hoc post-processing.

Finally, the hierarchy is designed to stay future-proof: reserved branches can support additional analog/RF/optical subsystems (e.g., optical filters, couplers/splitters, RF cable loss, and bandwidth effects) without altering the architectural logic. The main rule is that the naming convention follows “*block* → *property* → *mode/flag/value*,” ensuring that parameters remain human-readable, searchable, and consistent across blocks.

4.2.2 Enabling/disabling blocks and features

Because the surrogate is used for multiple research tasks (debugging, feasibility sweeps, calibration, and dataset generation), the configuration supports *feature gating*. Each major block and many sub-effects can be enabled/disabled through flags within the same “parameters” hierarchy. This avoids code branching scattered across files and keeps “what is active” visible in one place.

In practice, the framework supports the notion of simulation profiles: a profile is a coherent, preset configuration that activates a consistent subset of blocks and settings for a given goal. Typical examples include:

- **Fast-trend profile:** disables expensive dynamics and uses compact transfer-function abstractions to scan baud rate, bandwidth, and noise quickly.
- **Nominal waveform profile:** enables the complete waveform chain with realistic sampling, DAC/ADC effects, and measured/datasheet frequency responses for end-to-end feasibility evaluation.

- **Calibration profile:** enables measured-input hooks and calibration parameters (Section 4.2.4) while freezing irrelevant degrees of freedom to keep the identification problem well-posed.
- **ML-dataset profile:** keeps the physical chain fixed but activates systematic scenario generation (controlled sweeps over operating points) and structured metadata export.
- **Debug profile:** selectively bypasses blocks (e.g., channel or SOA) and reduces metadata collection to accelerate fault isolation.

Two implementation rules make these profiles scientifically useful. First, a profile must be explicitly declared (as a named configuration state), rather than inferred from manual edits across multiple files. Second, switching profiles should not change the mathematical meaning of a block; only its activation status and the parameter branch it uses should change, so that comparisons across profiles remain interpretable.

4.2.3 Reproducibility controls (random seeds, scenario IDs, sweep definitions)

Reproducibility is considered part of the configuration layer rather than an external “best effort.” In computational experiments, uncontrolled randomness (such as data shuffling, noise generation, Monte Carlo sampling, and randomized training splits) can lead to significantly different outcomes even when the high-level setup appears identical; therefore, seed control must be explicit and systematic^{19,20}. In the proposed framework, every stochastic component (e.g., bit generation, noise sources, phase-drift processes, randomized sweeps, and any future ML training routines) must use a declared seed policy rather than silently drawing from the default RNG state.

¹⁹ H. Ahmed and J. Lofstead, “*Managing Randomness to Enable Reproducible Machine Learning*”, Proceedings of the 5th International Workshop on Practical Reproducible Evaluation of Computer Systems, pp. 15-20, 2022.

²⁰ M. E. Schwamb, J. Kubica, M. Jurić, D. Oldag, M. West, M. DeLucchi, and M. J. Holman, “*Controlling Randomization in Astronomy Simulations*”, Research Notes of the AAS, vol. 8, no. 1, Article 25, 2024.

To make runs traceable at scale, each execution is associated with a *scenario ID*: a unique identifier that connects (i) the complete parameter snapshot, (ii) the seed policy, (iii) the enabled profile, and (iv) the produced outputs (waveforms, metrics, and metadata). This scenario ID is intended to be included in filenames and metadata exports so that results remain attributable after aggregation.

Parameter sweeps are treated as first-class objects instead of implicit loops in scripts. A sweep definition details the parameters to vary, their allowed ranges or sets, the stepping method (grid, randomized, or Latin hypercube if later needed), and how to handle seed repetitions (e.g., one seed for deterministic debugging versus multiple seeds for variance estimation). This structure enables fair comparisons across DSP/ML strategies by ensuring that different methods face identical scenario definitions and randomness controls.

Finally, reproducibility is improved by recording provenance: the framework is designed so that each run can be linked to the software/configuration state that created it (e.g., configuration dumps, scenario manifests, and version tags). This supports the broader view that computational results should be accompanied by a traceable record of the execution environment and configuration history, not merely the final figures²¹.

4.2.4 Calibration parameters vs nominal parameters (measured-input hooks)

A key architectural requirement of the digital surrogate is that each block can operate in two modes: a nominal mode, in which parameters reflect datasheet or expected values, and a calibrated mode, in which parameters are adjusted so that the block reproduces the measured behavior of the corresponding laboratory device at a specified operating point²². To make this distinction explicit and traceable, the global configuration layer separates (i) datasheet or nominal parameters (used to instantiate a physically plausible baseline) from (ii) calibration parameters (used to

²¹ A. Dhruv and A. Dubey, “*Managing Software Provenance to Enhance Reproducibility in Computational Research*”, Computing in Science & Engineering, vol. 25, no. 3, pp. 60-65, 2023.

²² P. D. Arendt, D. W. Apley, and W. Chen, “*Quantification of model uncertainty: Calibration, model discrepancy, and identifiability*”, Journal of Mechanical Design, vol. 134, no. 10, 100908, 2012.

fit the baseline to measurements), while preserving (iii) control or operation flags (which enable or disable features, select models, or choose resolutions), and (iv) knob parameters (degrees of freedom that correspond to controllable settings such as bias, gain, attenuation, filter selections, or operating point)²³.

The practical enabler of calibration is the availability of measured-input hooks: structured fields in the configuration that can store, reference, and version measurement-derived targets for each block, such as frequency responses, impulse/step responses, static transfer curves, noise PSD estimates, or waveform-level records captured at defined points of the chain²⁴. In optical link practice, hybrid modeling approaches (combining block simulation with measured responses or data-driven components) are widely used to obtain waveform-level predictions that remain tractable while matching real-system behavior across operating points²⁵. Within this thesis, “measured-input hooks” are treated as first-class objects: they are not *ad hoc* files loaded inconsistently, but standardized calibration inputs with identifiers (instrument settings, acquisition conditions, preprocessing steps), so that the same calibration target can be reused across studies and compared across model revisions²⁶.

4.2.4.1 Classical parameter fitting and identifiability

At the most direct level, calibration can be formulated as parameter estimation: select a subset of calibration parameters θ for a given block, define a mismatch metric $J(\theta)$ between measured outputs y_{meas} and simulated outputs $y_{\text{sim}}(\theta)$, and solve $\theta^* = \arg \min_{\theta} J(\theta)$ ²⁷. However, two recurring issues motivate more careful

²³ C. Li and S. Mahadevan, “*Role of calibration, validation, and relevance in multi-level uncertainty integration*”, Reliability Engineering & System Safety, vol. 148, pp. 32-43, 2016.

²⁴ *Ibidem*.

²⁵ H. Yang, Z. Niu, H. Zhao, S. Xiao, W. Hu, and L. Yi, “*Fast and Accurate Waveform Modeling of Long-Haul Multi-Channel Optical Fiber Transmission Using a Hybrid Model-Data Driven Scheme*”, Journal of Lightwave Technology, vol. 40, no. 14, pp. 4571-4580, 2022.

²⁶ P. D. Arendt, D. W. Apley, and W. Chen, “*Quantification of model uncertainty: Calibration, model discrepancy, and identifiability*”, Journal of Mechanical Design, vol. 134, no. 10, 100908, 2012.

²⁷ *Ibidem*.

treatment. First, *model discrepancy* (structural mismatch between the model class and reality) means that a perfect fit may not exist, and forcing a fit can produce non-physical parameter values²⁸. Second, *identifiability* limits how uniquely parameters can be inferred from available measurements; in many realistic settings, different parameter combinations yield similar outputs, especially when only a narrow set of metrics is used²⁹. For this reason, the surrogate’s configuration distinguishes between *what is being calibrated* (which parameters are allowed to vary) and *what is being matched* (which measurement targets are used), making calibration a repeatable process rather than an undocumented tuning exercise³⁰.

4.2.4.2 Bayesian calibration and uncertainty-aware fitting

A more robust alternative is *Bayesian calibration*, in which unknown calibration parameters are treated as random variables with a prior distribution $p(\theta)$, and measurements update this belief via a likelihood $p(y_{\text{meas}} | \theta)$ to produce a posterior $p(\theta | y_{\text{meas}}) \propto p(y_{\text{meas}} | \theta)p(\theta)$ ³¹. This approach is attractive for device-aware surrogates because it naturally supports (i) uncertainty in measurements, (ii) uncertainty in parameters, and (iii) explicit representation of model discrepancy (instead of hiding it inside tuned parameters)³². In practice, Bayesian calibration is often paired with compact discrepancy models (e.g., Gaussian-process discrepancy terms) to avoid attributing systematic mismatch entirely to physical parameters. For our purposes, the architectural implication is concrete: calibration parameters are stored together with priors/bounds and the selected likelihood/mismatch definition, so that “calibrated mode” is not merely a new constant set, but a documented inference result tied to specific measurement inputs and assumptions³³.

²⁸ M. C. Kennedy and A. O’Hagan, “*Bayesian calibration of computer models*”, Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 63, no. 3, pp. 425-464, 2001.

²⁹ *Ibidem* 8(P.D. Arendt *et al.*)

³⁰ C. Li and S. Mahadevan, “*Role of calibration, validation, and relevance in multi-level uncertainty integration*”, Reliability Engineering & System Safety, vol. 148, pp. 32-43, 2016.

³¹ *Ibidem* 10 (M.C. Kennedy *et al.*)

³² *Ibidem* 10 (M.C. Kennedy *et al.*)

³³ *Ibidem* 10 (M.C. Kennedy *et al.*)

4.2.4.3 Bayesian optimization as a practical calibration engine

Even when the calibration objective is not explicitly Bayesian, many calibration problems are *expensive black-box optimization* tasks because the surrogate may contain waveform-level simulation, memory effects, and nested blocks, making gradients unavailable or unreliable. In such cases, *Bayesian optimization (BO)* provides an efficient strategy: build a surrogate model (commonly a Gaussian process) for the objective $J(\theta)$, then iteratively select new parameter evaluations using an acquisition function that balances exploration and exploitation³⁴. BO has been demonstrated to be an effective, scalable tool for calibrating complex dynamical simulators against measurement data, particularly when each simulation run is costly³⁵. For the digital surrogate, this motivates storing calibration parameters with explicit bounds, selecting objective metrics that align with measurable quantities (e.g., transfer curves, response dynamics, spectrum/eye metrics), and logging the optimization trace so that the fitted result is reproducible and auditable rather than a one-off “good-looking” configuration³⁶.

4.2.4.4 ML-driven calibration and closed-loop tuning

In addition to Bayesian approaches, calibration can be performed via *learning-based strategies* that use data to accelerate or automate tuning. One route is supervised learning: learn an inverse mapping from observed features (measured waveforms/metrics) to parameters that best explain them, which can dramatically reduce calibration time once trained, but requires carefully curated training datasets and attention to identifiability³⁷. Another approach is *reinforcement learning (RL)*

³⁴ B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas, “*Taking the human out of the loop: A review of Bayesian optimization*”, Proceedings of the IEEE, vol. 104, no. 1, pp. 148-175, 2016.

³⁵ A. Chakrabarty, E. Maddalena, H. Qiao, and C. Laughman, “*Scalable Bayesian optimization for model calibration: Case study on coupled building and HVAC dynamics*”, Energy and Buildings, vol. 240, 111460, 2021.

³⁶ *Ibidem* 16 (B. Shahriari *et al.*)

³⁷ C. Li and S. Mahadevan, “*Role of calibration, validation, and relevance in multi-level uncertainty integration*”, Reliability Engineering & System Safety, vol. 148, pp. 32-43, 2016.

or policy-based optimization, in which an agent proposes parameter updates, receives a reward based on mismatch reduction, and learns a tuning policy that can operate online (or in repeated lab sessions) under changing conditions³⁸. The relevance to this thesis is not that a single method is “best,” but that the surrogate is built so all of them are feasible without restructuring code: calibration parameters, measurement hooks, objective definitions, and calibration logs are standardized at the configuration layer, enabling classical fitting (SysID), BO-based calibration, or ML/RL-based closed-loop tuning to be swapped as “calibration backends” while keeping the same block interfaces and metadata discipline.

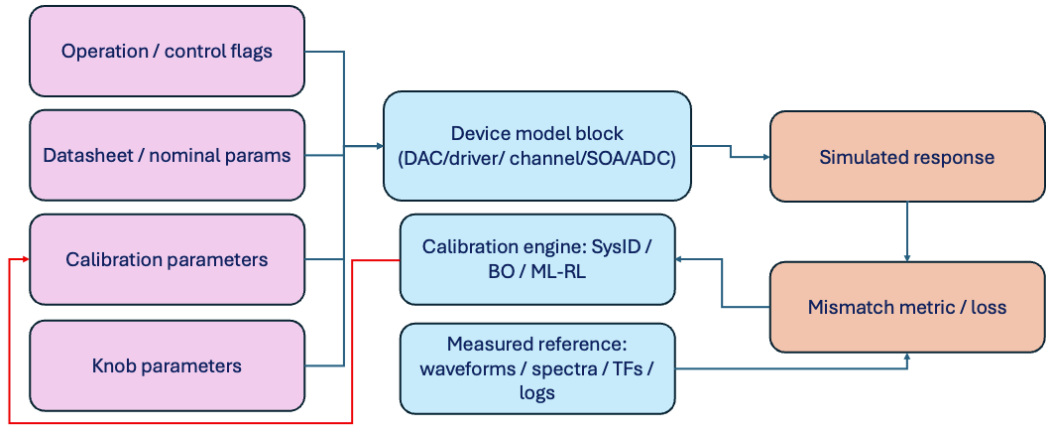


Figure 4.2 Calibration separation + measured-input hooks

4.3 Global runtime data structures

While Chapter 4 introduced the architectural principles and the global configuration layer, this section formalizes the *runtime* data structures that carry the evolving waveform and its context through the surrogate. The central idea is that an end-to-end surrogate is only scientifically usable if the signal state is transported in a structured, auditable form: the waveform itself, its sampling interpretation, its physical-domain meaning, and a disciplined record of which impairments have been applied and under what conditions. This aligns with established best practices in scientific computing, where clarity of data structures and traceability of

³⁸ H. M. Unjhwala, R. Zhang, W. Hu, J. Wu, R. Serban, and D. Negrut, “Using a Bayesian-Inference Approach to Calibrating Models for Simulation in Robotics”, Journal of Computational and Nonlinear Dynamics, 2023.

computational steps are considered essential for reproducibility and reliable comparison across experiments^{39,40}.

4.3.1 The signal structure

The signal (signal vector) structure is the primary runtime container passed between blocks. Its purpose is specifically twofold: (i) to carry the current waveform representation being transformed by the chain, and (ii) to maintain enough structured context to keep each transformation interpretable, reproducible, and calibration-ready.

The waveform container is “signal.current,” with the role as the canonical state. At every point in the chain, the waveform that represents “the current state of the link” is stored in “signal.current.” Each processing block reads from “signal.current”, applies its transformation, and writes the updated waveform back to “signal.current”. This makes the signal path explicit and enforces a single canonical handover point between blocks (rather than implicit dependencies on local variables or hidden shared state).

To assist with debugging and scientific ablation experiments, the structure can optionally store stage-specific snapshots of intermediate representations, such as generated bits, mapped symbols, upsampled sequences, pulse-shaped waveforms, and post-DAC electrical waveforms. These snapshots are controlled by the global configuration: they are useful for validation and reporting, but the framework does not depend on them for correctness.

Sampling information clarifies the discrete-time meaning. Because the surrogate operates at waveform (sample-level) resolution, each waveform must be interpreted in its sampling context. The structure, therefore, links the waveform with an explicit sampling rate field (e.g., “signal.Fs” in Hz), and the runtime

³⁹ G. Wilson, D. A. Aruliah, C. T. Brown, N. P. Chue Hong, M. Davis, R. T. Guy, S. H. D. Haddock, K. D. Huff, I. M. Mitchell, M. D. Plumbley, B. Waugh, E. P. White, and P. Wilson, “*Best Practices for Scientific Computing*”, PLOS Biology, vol. 12, no. 1, e1001745, 2014.

⁴⁰ C. Goble, S. Cohen-Boulakia, S. Soiland-Reyes, D. Garijo, Y. Gil, M. R. Crusoe, K. Peters, and D. Schober, “*FAIR Computational Workflows*”, Data Intelligence, vol. 2, no. 1-2, pp. 108-121, 2020.

pipeline ensures consistency among waveform length, sampling rate, and any derived timing assumptions needed by later blocks (such as resampling stages, filtering, or converter effects).

This avoids a common failure mode in end-to-end simulations: a waveform may be numerically valid but physically ambiguous if its sampling interpretation is not explicitly carried forward. Enforcing sampling metadata in the runtime container is also aligned with reproducibility practice: results must remain interpretable when re-run under different profiles or sweep settings⁴¹.

In this environment, domain tags explicitly indicate physical interpretations such as digital, electrical, or optical. A key property of IM-DD chains is that the waveform's *physical meaning changes* at specific boundaries: digital sequences, electrical drive waveforms, optical power or field representations, and detected electrical signals. Therefore, the “signal” structure includes a domain tag (e.g., “signal.domain”) that specifies how “signal.current” should be interpreted at each stage.

This is not just for appearance: it prevents silent category errors, such as applying an optical impairment model to an electrical waveform. Domain transitions are managed as *block responsibilities* (e.g., photodetection converts optical to electrical), so the domain tag offers a runtime assertion that the chain remains physically consistent.

In addition, lightweight stage tags (e.g., a block label and stage identifier) can be attached to the structure when needed, so that downstream tooling (plotting, validation, diagnostics) can interpret the waveform context deterministically using a lookup policy (as encoded in the visualization LUT used elsewhere in this chapter).

To track the impairment, structured, block-local metadata with summaries are collected. Besides transporting the waveform, the “signal” structure is designed to systematically monitor impairments. The guiding principle is that *impairments*

⁴¹ G. Wilson, D. A. Aruliah, C. T. Brown, N. P. Chue Hong, M. Davis, R. T. Guy, S. H. D. Haddock, K. D. Huff, I. M. Mitchell, M. D. Plumbley, B. Waugh, E. P. White, and P. Wilson, “*Best Practices for Scientific Computing*”, PLOS Biology, vol. 12, no. 1, e1001745, 2014

should be transparent and traceable to their injection point. In practice, each block can include a compact metadata record describing:

- **What was applied** (e.g., quantization, INL/DNL, jitter, bandwidth limitation, noise injection),
- **With which parameters** (a snapshot of the relevant configuration for that block),
- **The observable effect** includes summary statistics that are inexpensive yet informative: RMS levels, estimated SNR contributions, clipping rate, etc.

To keep this scalable, impairment records are stored in a nested namespace defined by block and stage (e.g., “signal.meta.<Block>.<Stage>”), rather than as unstructured logs. This method directly supports calibration workflows by aligning parameters and outputs at the correct injection point, and it also facilitates ML-ready dataset creation by linking each waveform to machine-readable provenance of operating conditions. The core motivation aligns with broader provenance and metadata guidelines: for computational models to be reusable and auditable, the data product must include sufficient structured context to reconstruct “how it was produced,” not just “what it is.”⁴²

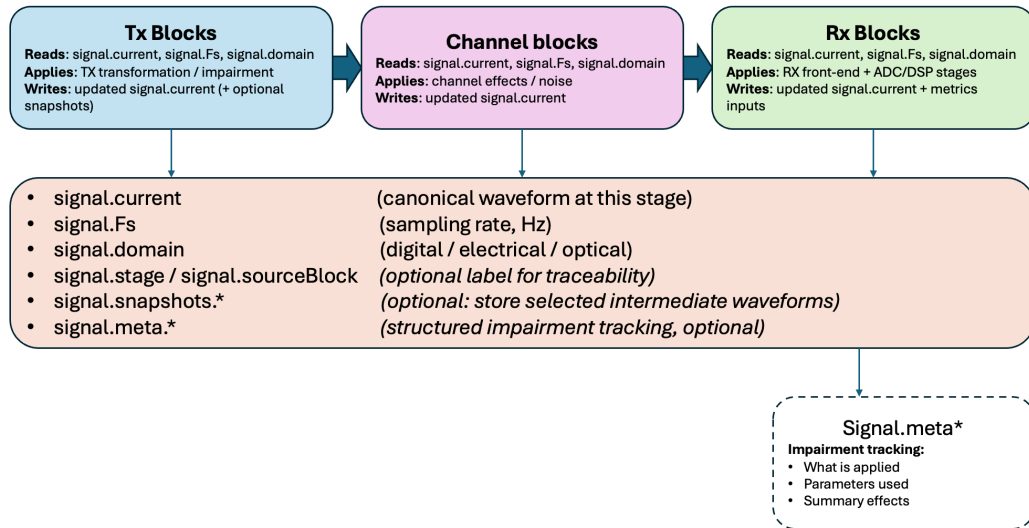


Figure 4.3 Global runtime data structure

⁴² C. Goble, S. Cohen-Boulakia, S. Soiland-Reyes, D. Garijo, Y. Gil, M. R. Crusoe, K. Peters, and D. Schober, “FAIR Computational Workflows”, Data Intelligence, vol. 2, no. 1-2, pp. 108-121, 2020.

4.3.2 The time structure and the time manager concept (time base, span, history)

In a sample-level surrogate, “time” must be regarded as a primary runtime object rather than an implicit index because almost all physically meaningful measurements (such as eye slicing, bandwidth-limited memory, jitter, SOA recovery, symbol timing, etc.) ultimately depend on a consistent mapping between sample index and physical time. Making this mapping explicit and verifying it after each transformation aligns with the general reproducibility principle, which states that hidden state should be minimized and that derived results should be traceable to well-defined, recorded computational conditions⁴³.

In this framework, the time structure acts as the primary carrier of that mapping. At a minimum, it offers a single authoritative time base throughout “time.current” (an $N \times 1$ times 1 vector aligned with “signal.current”) and the sampling descriptors “time.Fs” and “time.Ts = 1/time.Fs.” From these, it maintains lightweight yet useful summary fields: “time.n” (sample indices), “time.span” (total time span), and plotting helpers such as “time.units” (label + multiplier) and “time.axisLabel.” Additionally, it updates rate descriptors (“time.bitRate” from “parameters.tx.bitRate,” and “time.baudRate” derived from modulation parameters) so downstream instruments can reference the intended operating point without re-deriving it in each block.

The time manager (“time_manager(signal, time, parameters, callerTagOpt)”) is the enforcement mechanism that initializes, validates, and updates this structure as the waveform moves through the chain. It is intentionally simple in its contract: it only requires “signal.current” (vector) and “signal.Fs” (Hz); it returns the unchanged “signal” and an updated “time.” If “time” is empty, it can bootstrap the struct, but only at explicitly permitted stages (the default allow-list includes “generateBits”). If “time” exists, it performs consistency checks on two invariants: (i) length alignment (“length(time.current) == length(signal.current)”), and (ii) sampling consistency (“time.Fs == signal.Fs”). This form of defensive validation is

⁴³ G. K. Sandve, A. Nekrutenko, J. Taylor, and E. Hovig, “Ten Simple Rules for Reproducible Computational Research”, PLoS Computational Biology, vol. 9, no. 10, e1003285, 2013.

a recognized best practice in scientific software because it prevents silent, hard-to-diagnose errors from propagating into later metrics and conclusions⁴⁴.

A key design element is that the time manager supports policy-controlled handling of mismatches via the “parameters.timeManager.rebuildMode” parameter.

- “off”: don’t rebuild; flag inconsistencies (safest for strict debugging).
- “diagnostic”: generate *suggestions* for how time could be reconstructed (without implementing them), allowing traceable debugging.
- “apply”: rebuild “time.current” from the *current* “signal.Fs” and waveform length (useful when a known stage legitimately changes sample count or sampling rate and the pipeline should self-heal).

This makes “time correctness” explicit and configurable, rather than being an implicit assumption embedded in plotting or eye-diagram code.

Finally, the time manager maintains a bounded history (“time.history”) for auditing and debugging. Each update appends a compact snapshot (caller tag, previous caller, update type, key scalars like “Fs, Ts, N,” span, and a wall-clock timestamp), while enforcing a maximum length (default 200, “keepHistoryMax”) to prevent unbounded growth. This aligns with the broader scientific workflow concept that provenance should be recorded in a lightweight yet structured manner so results can be inspected and traced back to their origin steps⁴⁵. For easier downstream use, the manager also copies the authoritative time vector into the signal state (“signal.time = time.current”), allowing blocks and virtual instruments to depend on a consistent time base without tightly coupling to the internal time structure.

⁴⁴ G. Wilson, D. A. Aruliah, C. T. Brown, N. P. Chue Hong, M. Davis, R. T. Guy, S. H. D. Haddock, K. Huff, I. M. Mitchell, M. Plumbley, B. Waugh, E. P. White, and P. Wilson, “*Best Practices for Scientific Computing*”, PLoS Biology, vol. 12, no. 1, e1001745, 2014.

⁴⁵ J. Leipzig, D. Nüst, C. Tapley Hoyt, K. Ram, and J. Greenberg, “*The role of metadata in reproducible computational research*”, Patterns, vol. 2, no. 9, 100322, 2021.

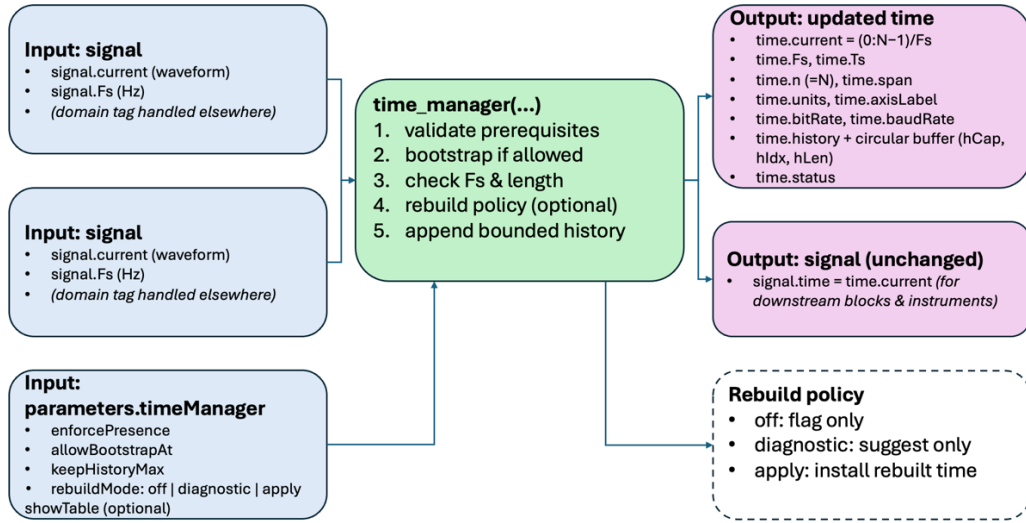


Figure 4.4 The time structure and the time manager concept (time base, span, history)

4.3.3 Visualization metadata (visMeta) and its separation from the main signal state

In an end-to-end IM-DD surrogate, visualization is not just an accessory; it is a crucial tool used to (i) verify that each block produces the intended physical transformation, (ii) identify performance issues at specific stages (e.g., “DAC nonlinearity vs. channel dispersion vs. AFE filtering”), and (iii) track the evolution of the waveform and its statistics throughout the chain. However, making visualization rely directly on the full “signal” state often leads to two practical problems: (1) plotting becomes fragile when internal representations change (e.g., sample count changes after ZOH or resampling), and (2) the model becomes more difficult to reproduce and extend because plotting logic begins to “reach into” block internals. For these reasons, the framework separates runtime signal state (“signal”) from visualization state (“visMeta”): “signal” remains the authoritative source of the physical waveform and impairment models, while “visMeta” provides a lightweight, standardized snapshot designed specifically for plotting and inspection.

This separation follows the broader reproducibility and provenance principle that computational pipelines should clearly display structured context around results (such as the stage that produced them and the conditions under which they

were generated), rather than depending on implicit state scattered throughout scripts⁴⁶.

Role and scope of visMeta. “visMeta” is viewed as a *representation* of the running system rather than the system itself. Specifically, the plotting entry point (“visualize_signal”) expects a “visMeta” structure that, at minimum, includes a “latest” snapshot of the waveform (“current”) along with the essential metadata needed for interpretation (sampling and stage context). If available, “visMeta” can also contain a “history” array of snapshots, allowing post-hoc inspection of intermediate stages without requiring the entire “signal” object to store heavy plotting-specific fields. This design keeps the simulation blocks clean (they update the physics) and maintains robust visualization (it reads from a stable, standardized schema).

What the snapshot contains (and why). Inspection of the plotting functions shows that the visualization layer consistently relies on a small set of “interpretation primitives” rather than block-specific internal variables. In practice, a “visMeta.latest” snapshot is intended to carry:

- “current”: the waveform vector to be visualized (real-valued for IM-DD electrical/intensity stages; extensible to complex for coherent fields later).
- “Fs” and/or a time axis: sampling frequency and time vector (“time”), allowing plots to be indexed in seconds instead of sample indices.
- “Rs” and “SpS” (when applicable): symbols rate and samples per symbol to enable UI-grids and eye folding.
- Stage identifiers: fields such as “stage,” “domain,” and the originating block tag (used for annotation and consistent plotting choices).
- Optional masks and markers: e.g., flags like an “original-sample mask” to differentiate raw discrete-time samples from ZOH-repeated samples. This distinction becomes important for eye construction and symbol alignment when the sample count varies.

This choice is deliberate: it guarantees that visualization remains consistent even during internal refactoring (the block may change its internal computation, but

⁴⁶ G. K. Sandve, A. Nekrutenko, J. Taylor, and E. Hovig, “Ten Simple Rules for Reproducible Computational Research”, PLOS Computational Biology, vol. 9, no. 10, e1003285, 2013.

the visualization contract stays the same). It also aligns with the common recommendation that computational results be accompanied by adequate structured context to support interpretation and reuse.

In this setup, the visualization lookup table acts as a “contract” layer. The function “generateVisualizationLUT.m” formalizes a key idea: the mapping between processing blocks and visualization semantics is declared once, in a central location. The LUT lists the entire process flow by stage (bit generation, mapping, upsampling, pulse shaping, transmitter DSP, DAC sub-stages, channel, SOA, photodetection, ADC sub-stages, receiver DSP, etc.) and assigns each stage a consistent set of tags (block name, domain, stage label, and comments). This prevents a common problem in large simulation codes: duplicated plotting and labeling logic that gradually diverges over time. Here, the LUT serves as the primary reference for “what stage are we in?” and “how should we label/interpret this waveform for visualization?” It also enables future automation: a collector can systematically build “visMeta” snapshots by reading the LUT rather than relying on manual bookkeeping for each block.

Our visualization entry point can be either a single dashboard or separate windows. The main rendering function (“visualize_signal”) supports two usage modes, chosen by configuration.

1. **Dashboard (2×2 layout):** a single figure hosting the four primary diagnostic views, time-domain, eye diagram, constellation view, and frequency spectrum, using shared styling and consistent stage context. This is the default “virtual instrument panel” for rapid iteration during development and debugging (Figure 4.5).

2. **Separate windows:** each plot can be opened independently. This is useful when exporting figures, when screen space is constrained, or when a particular view (e.g., an eye diagram) needs more resolution.

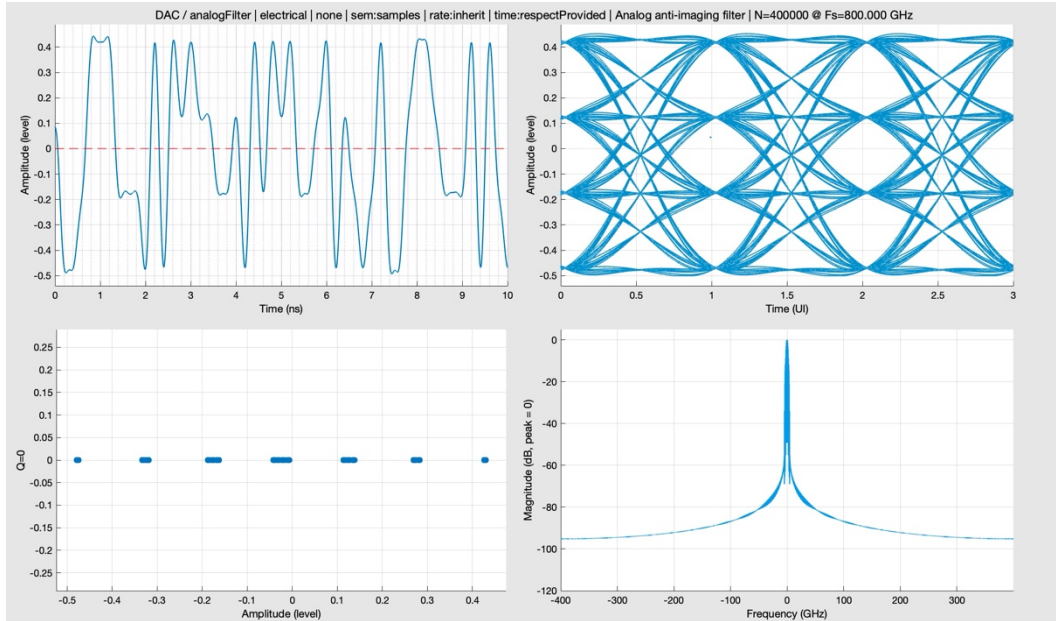


Figure 4.5 Visualization dashboard.

The dashboard mode also standardizes styling at the axes level (background, font, grid), while leaving plot-specific styles to each plotter. The code already includes explicit placeholders for consolidating styling into a single theme layer (e.g., to enforce a uniform publication look across all plots).

Plot engines and how they avoid sample-count pitfalls. The plotting layer is intentionally written to be **time-driven** whenever possible:

- **Time-domain (“plot_time_domain”)** prefers plotting against a time vector. It supports optional symbol grids when Rs/SpS are available, and it includes safeguards such as decimation for large arrays so plots remain readable and fast. Importantly, it can preserve interpretability across resampling/ZOH operations by relying on time rather than on raw sample index.
- **The eye diagram (“plot_eye_diagram”)** mainly uses folding in the time domain when a valid time base is available. This is exactly why the time-manager concept was created. After processes like ZOH repetition or downsampling, the relationship “samples-per-symbol” might not be reliable unless it is regularly recomputed. A time-driven fold (using the “time” axis and the symbol period) is more stable against such changes because it references the physics (seconds/UI) rather than array length. When time information is

incomplete, the function falls back to index-based folding in a controlled manner, but this is considered secondary.

- **Frequency domain** (“`plot_frequency_domain`”) computes spectra using the sampling frequency and includes guard logic for nonuniform time axes. It optionally reports bandwidth-oriented diagnostics (e.g., interpreting spectral content in a manner consistent with the configured sampling).
- **Constellation** (“`plot_constellation`”) supports sampling one value per symbol when symbol timing is known, which is useful even in PAM contexts (where “constellation” degenerates to 1D levels) because it reveals clustering, compression, and nonlinear warping of level distributions throughout the chain.

Several of these plot functions already include hooks for “scope-like” extensions, most clearly in the eye-diagram plotter, which contains optional metric annotations (eye opening/width proxies, crossing behavior, and related diagnostics). In the thesis, these should be framed as *diagnostic measurements* that can later be promoted to quantitative metrics (and possibly optimization objectives) once their definitions and calibration procedures are finalized.

It is crucial to understand that effective time management improves the accuracy of visualization. The time manager (“`time_manager.m`”) functions as the architectural safeguard that prevents visualization from depending on fragile assumptions about sample indices. It ensures consistency between “`signal.current`, `signal.Fs`,” and the corresponding time base; it can rebuild or diagnose mismatches under controlled policies; and it maintains a bounded history for traceability. When visualization uses a time-validated snapshot (rather than raw signal arrays), eye construction and UI-gridding stay clear even when blocks change sampling density (e.g., ZOH expansion) or when later stages apply downsampling or decimation. Consequently, visualization remains physically interpretable (“time in seconds / UI”) throughout the entire chain, rather than silently drifting into inconsistent indexing.

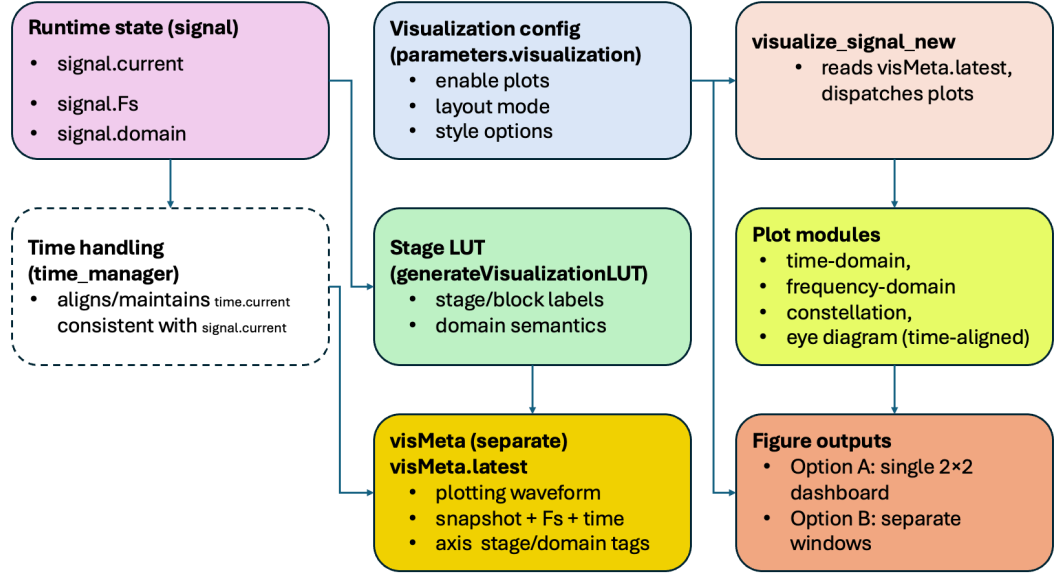


Figure 4.6 The visualization pipeline

4.3.4 ML/optimization metadata: what is stored, where it is stored, and why

A key design choice of the digital surrogate is that ML/optimization support is implemented not as a separate “ML module” attached to the simulator, but as a data discipline that accompanies the waveform throughout the chain. The reason is practical: if learning-based equalizers or predistortion models are trained solely on raw waveforms, without an explicit record of the operating point and active impairment mechanisms, the resulting models are challenging to reproduce, compare fairly, and often fail to generalize across conditions. This is especially true for IM-DD links, where performance depends on coupled mixed-signal effects (e.g., converter artifacts interacting with bandwidth limits and memory effects), and where the same measured BER/OSNR curve can result from different combinations of underlying impairments.

The stored metadata typically falls into specific categories. It is organized to facilitate three key tasks: (i) *supervised learning*, which includes labels and context; (ii) optimization or parameter sweeps, involving knobs and the objective; and (iii) *scientific interpretability*, focusing on attribution and ablation studies. In detail, this information is structured into:

- **Scenario and operating-point descriptors:** modulation format/order, bit/symbol rates, sampling rates, samples per symbol (when applicable), enabled simulation “profile,” random seed, and scenario ID, and any sweep

variables that define the run (e.g., OSNR/ASE setting, attenuation, SOA gain setting).

- **Block lineage and provenance:** which blocks were executed in the current run (and in what order), along with minimal provenance tags such as the block name/version or configuration hash when available (so that exported datasets can be traced back to an exact surrogate configuration).
- **Impairment parameters and realized statistics:** not only the “knob” values (e.g., DAC resolution, jitter RMS, INL/DNL settings, glitch pulse parameters), but also *the* summaries obtained during execution (e.g., estimated quantization step size, actual clipping fraction, measured noise power added by a block, or saturation indicators). This is important because ML models and optimization loops often require the *effective* impairment level, not just the nominal setting.
- **Supervised-learning labels aligned to the waveform:** transmitted bits/symbols (and optional intermediate representations such as Gray-mapped indices), timing/alignment references, and any task-specific targets. For BPAM with oversampled direct detection, the receiver explicitly extracts amplitude and phase information from two samples per symbol, thereby naturally defining label structures beyond “symbol value only.”⁴⁷ For CNN-based equalization/demapping, the learning task can be framed as amplitude/phase decoding (classification or regression), and the dataset must therefore store the target definitions and the segmentation/windowing rules used to generate training examples⁴⁸.
- **Optimization bookkeeping** involves tracking objective metrics (e.g., BER/EVM/OSNR penalty), constraint checks, and the knob settings for each evaluated candidate. This process enables later reconstruction of the optimization trajectory, including failures, and makes Bayesian optimization or RL-style calibration loops auditable rather than opaque.

⁴⁷ M. Secondini and E. Forestieri, “Direct Detection of Bipolar Pulse Amplitude Modulation”, *Journal of Lightwave Technology*, vol. 38, no. 21, pp. 5981-5990, 2020.

⁴⁸ R. Solaimani, A. Nizamani, S. Civelli, P. Nadimi Goki, F. Cavaliere, and L. Poti, “Experimental Demonstration of Convolutional Neural Network Equalization for BPAM in IM-DD Systems”, 2025.

The storage location (whether in-run or exported) is managed in the current system by storing ML/optimization metadata in the main runtime state, the same object that contains the waveform. Hierarchical fields are used so that each block writes to its own dedicated namespace, preventing overwrites of global data variables. In practice, this means: (i) *block-local metadata* is stored near the location where it is generated (e.g., DAC static/dynamic nonlinearity info; quantization stats; jitter stats), and (ii) *run-level metadata* is stored at a higher level (scenario ID, operating point, profile, sweep variable). The main policy is that metadata storage is controlled by gating (master enable + per-block enable), because full-fidelity metadata can become larger than the waveform itself during long sweeps. When export is needed, the surrogate writes a dataset record that includes waveform(s), labels, and the minimal metadata slice necessary to reproduce and interpret the record (rather than dumping the entire in-memory state).

The structure is based on the link to PAM/BPAM receiver goals. This design specifically aligns with the scientific aims of the thesis: it allows fair comparisons between DSP and ML approaches (using the same waveform, operating point, and impairment settings), supports controlled generalization experiments (training on one part of the operating space and testing on another), and enables meaningful ablation studies (for example, removing only INL/DNL, disabling SOA memory, or changing only the sampling policy). It also aligns with the needs of BPAM IM-DD receivers, where feasibility depends on time resolution and how transceiver impairments distort the oversampled structure used for reconstruction. When ML equalizers are used to compensate transceiver impairments in BPAM IM-DD experiments, the learning pipeline (windowing, label definition, train/test split, and performance reporting) becomes part of the scientific method. It should be generated from stored metadata rather than reconstructed later from other sources memory. Finally, the same schema supports evolution: as the project expands to additional components or coherent variants, the metadata remains the stable backbone, with only block-specific namespaces growing.⁴⁹

⁴⁹ *Ibidem*.

4.3.5 Logging and diagnostics conventions (verbosity, warnings, failure modes)

In the current implementation, diagnostic output mainly functions as an interactive debugging tool. Each block can display brief status messages when it is “in focus,” while non-central blocks are typically silenced to prevent flooding the command window. This method aligns with the role of logs in reliability engineering, as runtime logs are often the only clues to what the system actually did. Their usefulness relies heavily on clarity, consistency, and proper context information⁵⁰.

The verbosity policy (human-readable diagnostics) is structured in layers: (i) a master switch for global on/off control, and (ii) block-level toggles that activate only when needed. This prevents overwhelming outputs while still allowing detailed inspection during diagnostics, such as identifying a DAC nonlinearity combined with filtering and sampling alignment issues. In software engineering studies of logging practices, this “selective visibility” principle is repeatedly stressed: logs are useful, but uncontrolled verbosity generates noise, obscures signals, and hampers troubleshooting⁵¹.

Warnings versus errors (failure-mode discipline) that we needed to address, to ensure debugging remains scientifically meaningful, the surrogate differentiates between:

- **Hard failures (errors):** Violations that invalidate results (e.g., missing mandatory fields, domain mismatch, impossible sampling configuration). These should halt execution and provide a clear identifier and message that indicate the responsible block and the violated precondition.
- **Soft failures (warnings):** non-fatal conditions that may still allow progress but should be visible (e.g., a parameter value is outside a

⁵⁰ S. He, P. He, Z. Chen, T. Yang, Y. Su, and M. R. Lyu, “A Survey on Automated Log Analysis for Reliability Engineering”, ACM Computing Surveys, vol. 54, no. 6, art. 130, pp. 1-37, 2021

⁵¹ S. Gu, G. Rong, H. Zhang, and H. Shen, “Logging Practices in Software Engineering: A Systematic Mapping Study”, IEEE Transactions on Software Engineering, vol. 49, no. 2, pp. 902-923, 2022.

recommended range; a calibration hook is enabled, but required measured inputs are missing, causing the code to revert to nominal behavior).

This separation supports interpretability: if a result is produced after a warning, the warning becomes part of the experimental record. At scale, these conventions also make automated log processing feasible (e.g., filtering by severity, clustering repeated warning types), a central focus of the log-analysis literature⁵².

As a recommended improvement, we should continue logging as a reproducibility artifact. While the current workflow depends on console output, it is advantageous to *save logs optionally*, not as a substitute for interactive debugging, but as a reproducibility and calibration resource. The best practice is:

1. **A run identifier** (scenario ID) created at the beginning of execution.
2. **A parameter snapshot** stored alongside outputs (or a hash pointing to it).
3. **A log file** saved per run, containing time-stamped entries with: “{level, blockTag, message, key parameters}”.

This design supports two key workflows: (i) calibration runs, where you later need to trace why an optimizer or RL loop converged or diverged under a specific measurement set; and (ii) ML dataset generation, where you must ensure traceability from a training batch back to the exact impairment configuration. The importance of metadata and provenance in making computational research reusable and auditable is well recognized, and logging becomes part of that provenance layer when combined with configuration snapshots and identifiers⁵³.

For better scalability and organization, if logs are stored, the framework should prioritize using structured messages (including minimally structured “key=value” fields) over purely free-form text. This reduces ambiguity later and supports log abstraction/parsing as run volumes grow, an issue that becomes unavoidable as experiments scale to sweeps and dataset generation⁵⁴.

⁵² *Ibidem* (S. He *et al.*)

⁵³ G. K. Sandve, A. Nekrutenko, J. Taylor, and E. Hovig, “Ten Simple Rules for Reproducible Computational Research”, *PLoS Computational Biology*, vol. 9, no. 10, e1003285, 2013.

⁵⁴ D. El-Masri, F. Petrillo, Y.-G. Guéhéneuc, A. Hamou-Lhadj, and A. Bouziane, “A Systematic Literature Review on Automated Log Abstraction Techniques”, *Information and Software Technology*, vol. 122, art. 106276, 2020.

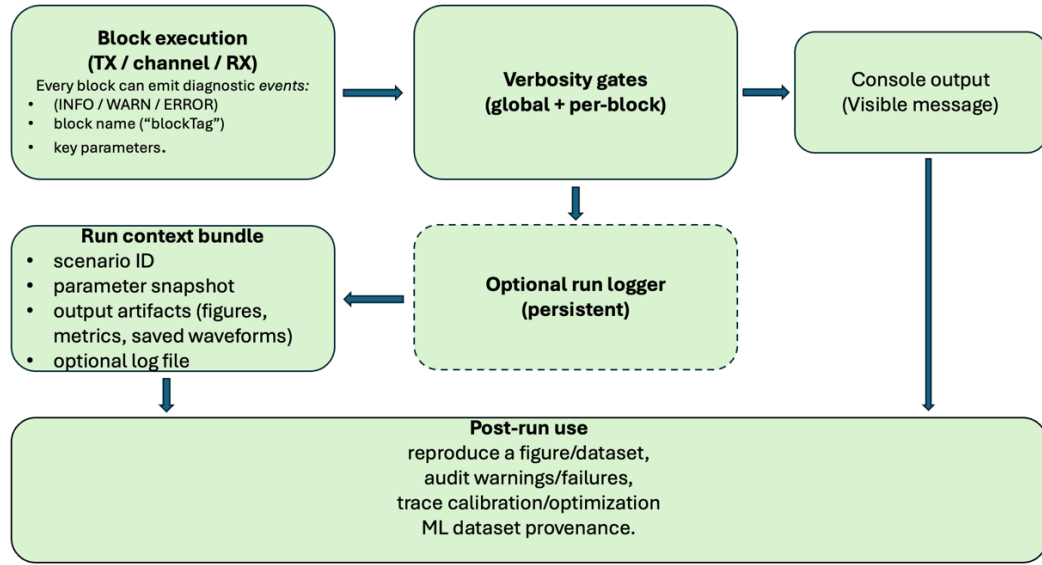


Figure 4.7 Logging and diagnostics flow in the digital surrogate

4.4 Block interface conventions and execution model

4.4.1 End-to-end chain overview

The surrogate operates as a deterministic block pipeline, where each stage performs a specific physical or algorithmic transformation and exposes its controllable parameters through the global “parameters” configuration. This execution model reflects how IM-DD link studies typically separate (i) digital transmitter processing, (ii) waveform generation and analog impairments, and (iii) receiver-side sampling and DSP evaluation, facilitating consistent benchmarking of BER, EVM, and eye-diagram analysis across various operating conditions points⁵⁵.

In our setup, the transmitter begins with bit generation or import (including reproducible RNG control), symbol mapping (e.g., Gray coding and standardized PAM alphabets), and waveform-level DSP steps such as upsampling, pulse shaping, and optional pre-emphasis or pre-distortion. The waveform then passes through the DAC parent block, which calls device-internal sub-functions (e.g., sample-rate handling, remapping, quantization, static nonlinearities such as INL/DNL, ZOH,

⁵⁵ T. Wettlin, S. Calabrò, T. Rahman, J. Wei, N. Stojanovic, and S. Pachnicke, “*DSP for High-Speed Short-Reach IM/DD Systems Using PAM*”, *Journal of Lightwave Technology*, vol. 38, no. 24, pp. 6771-6778, 2020.

and additional noise or impairment stages when enabled). After the DAC, the chain continues through driver and electrical conditioning, electro-optic modulation, fiber and channel effects, and optional optical amplification (SOA pre-amplifier), followed by photodetection and analog front-end conditioning (PD/TIA/AFE), ADC sampling and quantization, and finally receiver DSP, equalization, and metric extraction.

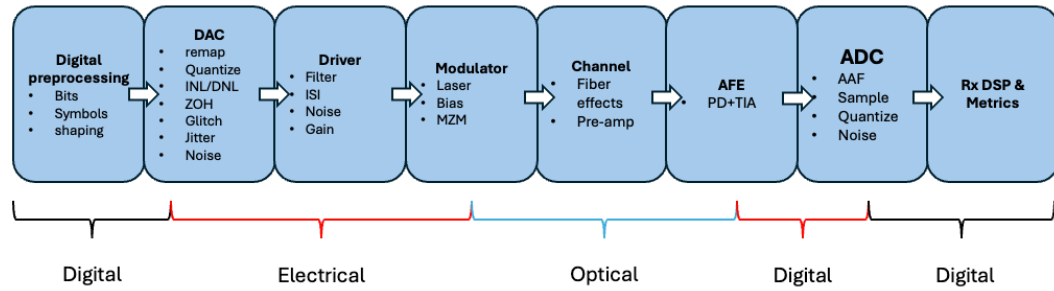


Figure 4.8 End-to-end execution chain and data hand-offs

4.4.2 Standard block I/O contract (inputs, outputs, required fields, optional fields)

To keep the chain modular and calibration-ready, each block follows a stable “I/O contract”: the block receives a runtime state (waveform + context), applies a transformation controlled only by “parameters,” and returns an updated state. This pattern aligns with established best practices for scientific software, where maintainability and correctness improve when modules have clear responsibilities, explicit inputs/outputs, and minimal hidden state⁵⁶.

Required inputs (conceptual):

- “signal.current”: the waveform vector currently carried through the chain (domain depends on stage).
- “signal.Fs”: sampling rate in Hz for the current waveform representation.
- “signal.domain”: an explicit domain tag (e.g., digital/electrical/optical) used to enforce correct placement of impairments and units.
- “parameters”: the global configuration structure containing all block settings and switches.

Optional inputs/side-structures (enabled by flags):

⁵⁶ G. K. Sandve, A. Nekrutenko, J. Taylor, and E. Hovig, “Ten Simple Rules for Reproducible Computational Research”, *PLOS Computational Biology*, vol. 9, no. 10, e1003285, 2013.

- “time”: a time structure used to preserve an unambiguous time base even when sample counts change (e.g., after ZOH replication or resampling).
- “visMeta”: visualization metadata describing what the waveform represents and how to plot/interpret it.
- “mlMeta” (or an equivalent metadata store): structured metadata for ML/optimization dataset generation.

Outputs (conceptual):

- Updated “signal” (and domain tag if a domain transition occurred).
- Updated “time” (when time management is enabled/needed).
- Updated “visMeta” and/or ML metadata when their corresponding switches are enabled.

Critically, blocks must not “reach into” other blocks’ parameters or hardcode defaults: if a value affects results (e.g., quantizer resolution, bandwidth corner, noise bandwidth), it must be exposed in “parameters” so that runs are traceable, transparent, and reproducible⁵⁷.

4.4.3 Unit conventions and domain transitions (what must be explicit at each boundary)

Because IM-DD chains combine digital arithmetic, electrical waveforms, and optical power or field quantities, unit errors are among the quickest ways to produce convincing yet incorrect simulation results. The execution model, therefore, treats *domain transitions* as explicit boundaries: whenever the signal shifts from digital to electrical, electrical to optical, optical back to electrical, or electrical to sampled digital, the block handling the transition must declare (and, where appropriate, store) the assumed physical meaning and scaling.

Practically, this means:

- **Digital domain:** discrete-time sample sequences with dimensionless amplitudes (unless explicitly normalized to a physical scale).
- **Electrical domain:** voltages/currents with stated conventions (e.g., V into 50 Ω ; bandwidth models expressed in Hz).

⁵⁷ G. Wilson, D. A. Aruliah, C. T. Brown, N. P. Chue Hong, M. Davis, R. T. Guy, S. H. D. Haddock, K. D. Huff, I. M. Mitchell, M. D. Plumbley, B. Waugh, E. P. White, and P. Wilson, “*Best Practices for Scientific Computing*”, *PLOS Biology*, vol. 12, no. 1, e1001745, 2014.

- **Optical domain:** optical intensity/power or complex field (depending on the chosen abstraction level), with noise sources applied only where physically meaningful.

- **AFE boundary:** the analog front end (AFE) is treated as the analog conditioning chain following the photodiode (including the option of using the transimpedance amplification and filtering), and its bandwidth/noise must not be “silently” merged into unrelated blocks.

More generally, mature scientific software ecosystems prioritize explicit handling of quantities/units and time because they prevent hidden inconsistencies and simplify transformations that are auditable⁵⁸.

4.4.4 How blocks update signal, time, and optional metadata consistently

Each block updates the runtime state according to a consistent rule set:

1. *Read* “signal.current” and the minimal required context (“signal.Fs”, “signal.domain”, and the block’s parameter subtree).
2. *Validate* prerequisites (presence of required fields; domain compatibility; basic range checks).
3. *Transform* the waveform (deterministic given parameters + RNG seed settings).
4. *Write back*:
 - “signal.current” updated to the new waveform,
 - “signal.domain” updated if a domain transition occurred,
 - “signal.Fs” updated if sampling rate changed,
 - optional impairment tracking fields updated if enabled (e.g., storing what was applied and with which parameters).
5. *Update time* through the time manager when enabled so that plots and measurements remain time-consistent even if sample length changes (e.g., ZOH replication, resampling, decimation).
6. *Attach metadata* only when requested:

⁵⁸ T. P. Robitaille et al. (Astropy Collaboration), “*Astropy: A community Python package for astronomy*”, *Astronomy & Astrophysics*, vol. 558, A33, 2013

- “visMeta” for visualization traceability,
- ML/optimization metadata for dataset lineage, operating-point labels, and training-supervision targets.

This “always update the state the same way” discipline is what makes large parameter sweeps and ML dataset generation reliable: the run becomes a reproducible execution of a declared configuration rather than a collection of implicit side effects⁵⁹.

4.4.5 Lightweight execution modes (debug vs full-fidelity vs ML-data-gathering modes)

A practical surrogate must support multiple *execution profiles* without requiring code edits. Therefore, the framework distinguishes at least three modes:

- **Debug mode:** minimal blocks enabled, simplified visualization, verbose logs on the block of interest, and (optionally) shortened waveform lengths for faster iteration.
- **Full-fidelity mode:** fully enabled impairment chain with physically meaningful parameters and time-consistent visualization.
- **ML-data-gathering mode:** uses the same physics as full-fidelity but with structured metadata export enabled, stable scenario identifiers, and strict reproducibility controls (seeds, run IDs, and configuration capture) so datasets are traceable and comparable.

This approach follows well-established reproducibility guidance: record how results were produced (including randomness control), keep configurations externalized, and separate “fast iteration” from “final data generation” without changing the scientific meaning of the pipeline⁶⁰.

⁵⁹ G. Wilson, D. A. Aruliah, C. T. Brown, N. P. Chue Hong, M. Davis, R. T. Guy, S. H. D. Haddock, K. D. Huff, I. M. Mitchell, M. D. Plumbley, B. Waugh, E. P. White, and P. Wilson, “*Best Practices for Scientific Computing*”, *PLOS Biology*, vol. 12, no. 1, e1001745, 2014.

⁶⁰ G. K. Sandve, A. Nekrutenko, J. Taylor, and E. Hovig, “*Ten Simple Rules for Reproducible Computational Research*”, *PLOS Computational Biology*, vol. 9, no. 10, e1003285, 2013.

4.5 Visualization and “virtual instruments”

A digital surrogate's credibility depends on its ability to reveal internal states in ways that mirror what engineers would observe in a laboratory setting. In practice, optical-communication experiments are analyzed using instruments such as oscilloscopes for time-domain and eye analysis, optical spectrum analyzers (OSAs) for spectral and OSNR measurements, and constellation/EVM tools after DSP. Therefore, the surrogate views visualization not as “pretty plotting” but as a virtual instrumentation layer that (i) makes intermediate signals accessible at each block boundary, (ii) extracts consistent measurements from sampled waveforms, and (iii) generates repeatable, thesis-ready figures through a controlled export pipeline.

4.5.1 Rationale: virtual instrumentation as part of surrogate credibility

Virtual instruments serve two primary purposes. First, they offer *scientific interpretability*: when performance varies, users must be able to trace the change to a specific stage (e.g., DAC quantization, driver filtering, MZM biasing, SOA saturation, ADC nonidealities), rather than just seeing an overall BER/EVM value. Second, they enable *comparability*: the surrogate outputs should be analyzable using the same “measurement logic” employed in the lab (eye metrics, spectral noise-band reasoning, timing alignment), so that calibration between simulation and experiment remains meaningful.

This view aligns with how modern IM-DD monitoring and impairment diagnosis are often described: eye diagrams are treated as measurement objects, generated by oscilloscope-like modules and then analyzed to determine eye parameters and quality metrics⁶¹. Likewise, eye-based alignment and clock recovery methods can link waveform sampling, timing, and jitter-related observations to performance trends in high-speed optical signals⁶².

⁶¹ D. Wang, Y. Xu, J. Li, M. Zhang, J. Li, J. Qin, “Comprehensive Eye Diagram Analysis: A Transfer Learning Approach”, IEEE Photonics Journal, vol. 11, no. 6, Art. no. 7905019, 2019.

⁶² S. Liu, “Research on high-speed digital optical signal jitter measurement technology based on clock recovery algorithm using eye diagram opening area”, Heliyon, vol. 10, Art. e35756, 2024.

4.5.2 Time-domain visualization and summary measurements

Time-domain plots are the primary “first look” tool in the surrogate because they reveal waveform scaling, clipping, filtering, memory, and noise injection most directly. In the current framework, time-domain visualization is considered *measurement-ready* rather than purely qualitative: each plotted segment can include a brief set of summary observables, such as mean value and DC offset, RMS level, peak level, crest factor, and optionally level statistics for multi-level PAM/bPAM. These quantities are not meant to replace end-to-end metrics (BER/EVM), but to act as quick diagnostics that identify faults early (e.g., incorrect normalization before DAC, unexpected bias drift, or bandwidth mismatch).

A key design requirement is correct time-base handling, even when the number of samples changes due to operations like zero-order hold (ZOH), resampling, or downsampling. Many simulation codes rely on “samples-per-symbol” for eye construction and time plots, but this becomes fragile once sample-rate-changing blocks are added. Therefore, the surrogate uses a consistent runtime, time base (via the time manager concept introduced in Section 4.3.2) so that visual instruments operate on time instead of fragile sample indexing. This approach also aligns with the needs of waveform-based estimators for optical intensity links, where timing and observation windows, along with sampling consistency, directly influence the reliability of extracted SNR-related measurements⁶³.

4.5.3 Eye diagram engine: scope-like behavior and alignment strategy

The eye diagram is the surrogate’s most important virtual instrument for IM-DD credibility because it visually aggregates intersymbol interference, bandwidth limitation, nonlinear distortion, and timing uncertainty into a single diagnostic view. The framework, therefore, implements an *eye diagram engine from scratch* (rather than relying on MATLAB’s built-in plotting), specifically to (i) control alignment assumptions explicitly, (ii) ensure consistent behavior after sample-rate-

⁶³ W. Gappmair, “Non-Data-Aided SNR Estimation for Bandlimited Optical Intensity Channels”, *Sensors*, vol. 23, no. 2, Art. 802, 2023.

altering operations, and (iii) enable direct extraction of standard eye metrics within the same implementation.

Scope-like construction: Conceptually, the engine follows the same principle used by oscilloscope-hosted eye generators: it converts a synchronously sampled waveform into an eye by segmenting the waveform into repeated windows and overlaying them on a common time axis. This is consistent with IM-DD studies that explicitly describe “oscilloscope-hosted eye-diagram generation modules” and then extract eye parameters such as eye height, eye width, crossing percentage, and Q-related indicators from the generated eye representation⁶⁴.

Alignment strategy: Eye construction is only meaningful if the overlay is aligned to a symbol clock or a recovered timing reference. The surrogate supports a time-manager-assisted alignment strategy: it anchors segmentation in the global time base rather than raw sample indices and aligns segment boundaries to the expected UI timing implied by the configured symbol rate. When necessary, such as in cases of timing drift or jitter-like effects, the alignment logic can be extended to clock-recovery-inspired approaches, in which the alignment is chosen to maximize an eye-opening criterion. The importance of eye-opening-area concepts for clock recovery and jitter measurement in high-speed optical signals has been explicitly discussed in recent literature⁶⁵.

Standard eye metrics: The engine is designed so that pair eye visualization with extractable metrics (eye height/width, crossing-related observables, and jitter-related proxies, where applicable). This reflects a broader practice in IM-DD monitoring: the eye is not only a picture, but also a quantitative measurement object used for performance monitoring and diagnosis⁶⁶.

⁶⁴ *Ibidem*.

⁶⁵ S. Liu, “Research on high-speed digital optical signal jitter measurement technology based on clock recovery algorithm using eye diagram opening area”, *Heliyon*, vol. 10, Art. e35756, 2024.

⁶⁶ D. Wang, Y. Xu, J. Li, M. Zhang, J. Li, J. Qin, “Comprehensive Eye Diagram Analysis: A Transfer Learning Approach”, *IEEE Photonics Journal*, vol. 11, no. 6, Art. no. 7905019, 2019.

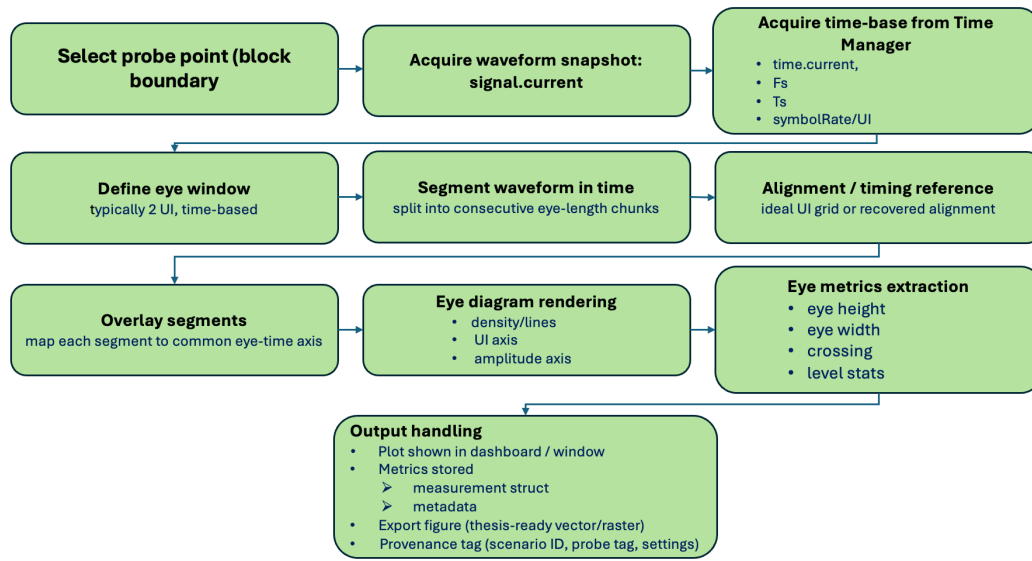


Figure 4.9 Virtual oscilloscope pipeline for eye formation and metrics.

4.5.4 Spectral tools and OSNR/SNR-related observables

The surrogate’s frequency-domain instrument is designed to emulate the reasoning used in OSA and digital spectral estimators. Its goal is not only to display spectra but also to support *OSNR/SNR-related observables* that are directly applicable to experimental practice.

OSA-like OSNR reasoning (ASE interpolation): In operational optical systems, OSNR is commonly estimated from spectra by identifying the signal band and estimating ASE noise from neighboring spectral regions (followed by interpolation into the signal band under defined assumptions). This OSA-centered perspective, spectral processing, carrier-frequency alignment, and mapping of spectral features to OSNR have been explicitly described in an in-band OSNR estimation workflow based on measured optical spectra⁶⁷. In the surrogate, this motivates implementing spectrum displays and band-power reporting that are compatible with “noise-band” reasoning, rather than simply plotting an FFT magnitude.

⁶⁷ F. Locatelli, K. Christodouloupoulos, M. Svaluto Moreolo, J. M. Fàbrega, and S. Spadaro, “Machine Learning-Based In-Band OSNR Estimation From Optical Spectra”, IEEE Photonics Technology Letters, vol. 31, no. 24, pp. 1929-1932, 2019.

Waveform-based SNR estimation (time-domain observables):

Complementary to OSNR reasoning, the surrogate also supports time-domain SNR observables derived directly from sampled intensity waveforms, which is particularly relevant when the receiver access point is electrical (post-PD/TIA, post-ADC) and spectral OSNR estimation is not the appropriate instrument. Non-data-aided SNR estimation methods for bandlimited optical intensity channels highlight the practical dependence of SNR estimation on timing assumptions and observation window selection⁶⁸. Recent extensions propose architectures that separate noise-dominated samples via filtering strategies to improve estimation robustness and efficiency, thereby motivating the design of the surrogate’s measurement blocks to optionally expose “noise-only” observables when such structures are implemented⁶⁹.

4.5.5 Export rules for figures and repeatable plots

Because thesis figures must be regenerated reliably throughout writing and revision, the surrogate treats figure generation as part of a *repeatable analysis pipeline* rather than an interactive, manual plotting task. Two best-practice ideas guide this:

1. **Workflow-encoded plotting pipelines.** Reproducible research workflows emphasize that plots should be regenerable from raw data via scripted pipelines with explicit parameters, rather than relying on undocumented manual steps⁷⁰. In the surrogate context, this motivates: fixed plot “templates” (styles, axes labels/units, window sizes), deterministic naming rules for outputs, and a consistent mapping from a simulation scenario ID to exported figures.

2. **Provenance of plot settings and inputs.** Provenance-capture approaches show that figure outputs are most valuable when accompanied

⁶⁸ W. Gappmair, “Non-Data-Aided SNR Estimation for Bandlimited Optical Intensity Channels”, *Sensors*, vol. 23, no. 2, Art. 802, 2023.

⁶⁹ W. Gappmair, “Novel Results on SNR Estimation for Bandlimited Optical Intensity Channels”, *Sensors*, vol. 24, no. 1, Art. 23, 2024.

⁷⁰ F. Mölder *et al.*, “Sustainable data analysis with Snakemake”, *F1000Research*, vol. 10, Art. 33, 2021.

by machine-readable records of inputs, parameter settings, and processing steps, enabling later audit and reuse⁷¹. In the surrogate, this motivates storing (at minimum) the scenario identifier, the parameter snapshot used to generate the plot, and the signal lineage tag indicating *where in the chain* the plotted waveform was taken.

In practice, the export rules are therefore: (i) figures are saved through a controlled export function, (ii) vector formats are used where possible for publication quality (e.g., PDF/SVG), and (iii) a small “figure manifest” (scenario ID, plot type, block tag, timestamp, and parameter hash) is generated so that each figure is traceable to an exact run configuration.

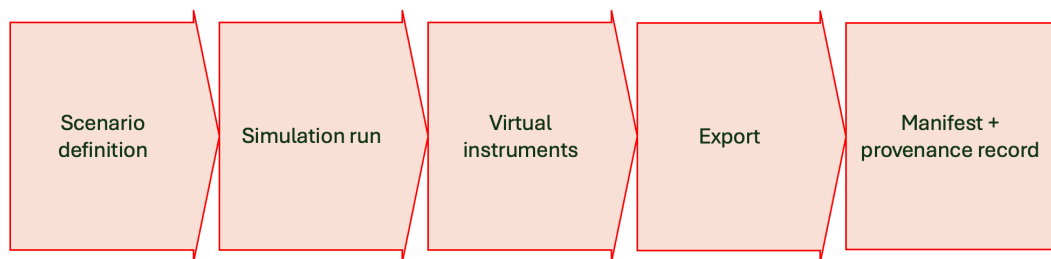


Figure 4.10 Reproducible figure export and provenance.

4.6 Internal consistency checks and validation strategy

Because the digital surrogate is designed to be reusable, calibratable, and ML-ready, it must exhibit predictable behavior under both normal and “stress” operating conditions. This section outlines the internal consistency checks and validation strategies used to maintain the framework's scientific credibility as model fidelity increases and as blocks are swapped, extended, or calibrated. The main goal is to ensure correctness at multiple levels: (i) global time-base consistency, (ii) unit/range/configuration sanity, (iii) block-level validation of component behavior, and (iv) end-to-end regression tests to identify unintended changes across code revisions.

⁷¹ C. A. Köhler *et al.*, “Facilitating the Sharing of Electrophysiology Data Analysis Results Through In-Depth Provenance Capture”, eNeuro, 2024.

4.6.1 Time-base consistency checks (Fs/length mismatches, rebuild policies)

A key reliability risk in sampled-waveform simulations is silent time-base drift: the waveform length and the assumed sampling rate can become inconsistent after operations that alter the sample count (e.g., upsampling, ZOH repetition, downsampling, trimming, buffering). To address this, the framework manages the time base as an explicitly controlled runtime object (the time structure), updated through a single gatekeeper (the time manager). The time manager maintains a “single source of truth” for sampling rate (F_s), sampling period (T_s), sample indices (n), and span, and detects mismatches as soon as they occur appear.

The validation policy operates in a rebuild mode with three practical behaviors. In *off* mode, inconsistencies are detected and reported but not automatically fixed (useful for debugging errors). In *diagnostic* mode, the time manager suggests repairs (e.g., regenerating *time.current* to match the updated F_s or the new vector length) while keeping the current time base unchanged, allowing the user to verify the cause of the mismatch. In *apply* mode, repairs are performed automatically and logged, supporting robust “production” runs where different blocks might change the waveform length. At the same time, the visualization and measurement pipeline must stay stable.

This style of explicit, global consistency enforcement follows the same engineering logic used in collaborative model-based workflows: consistency checks must be systematic, always-on, and capable of detecting cross-artifact contradictions early, rather than relying on manual inspection⁷².

4.6.2 Range checks, unit sanity checks, and configuration validation

Beyond timing, simulation credibility relies on preventing unit and range errors from spreading through the chain (often without obvious runtime failures). Therefore, the framework applies three categories of sanity checks:

1. **Range checks (numerical sanity):** Ensure signals contain finite values (no NaN/Inf) and that expected bounds are respected (e.g., normalized DAC input ranges before remapping, clipping policies, non-

⁷² M. A. Tröls, L. Marchezan, A. Mashkoo, and A. Egyed, "Instant and global consistency checking during collaborative engineering," *Software and Systems Modeling*, vol. 21, pp. 2489-2515, 2022.

negative intensity constraints where applicable). When a bound is violated, the framework reports the responsible block and relevant parameter settings to facilitate troubleshooting actionable.

2. **Unit and domain sanity:** At each domain transition (digital → electrical → optical → electrical), the interface contract must clearly specify what the waveform represents (voltage/current/field/intensity), which scaling convention is used (linear amplitude vs. power-normalized), and which noise model is physically relevant in that domain. This is not just a software hygiene issue; it is a scientific requirement because applying an impairment in the wrong domain can produce visually convincing but physically invalid results. The general principle aligns with “physical validity” testing practices in other computational physics fields: a model should be verified against basic physical invariants and sanity constraints, not only numerical correctness⁷³.

3. **Configuration validation:** Before execution, the framework verifies that required configuration fields exist, that enumerated options are valid (e.g., selected modulation type, enabled noise types), and that parameter combinations are consistent (e.g., symbol rate compatible with sampling rate and chosen samples-per-symbol; bandwidth parameters positive and plausible). This prevents “half-configured” runs that silently revert to defaults or execute with unintended assumptions.

4.6.3 Block-level validation (unit tests/component sanity tests)

Internal consistency must also be demonstrated locally: each block should be testable as a scientific component, not only as part of the entire chain. Therefore, the validation strategy includes block-level sanity tests that can be run with minimal dependencies. Typical examples include:

⁷³ P. T. Merz and M. R. Shirts, "Testing for physical validity in molecular simulations," PLoS ONE, vol. 13, no. 7, e0202764, 2018.

- **Deterministic reference tests:** A fixed input and seed produce exactly reproducible output (bit generation, PRBS generation, mapping, deterministic filtering).
- **Analytical spot checks:** When a closed-form expectation is available (such as energy change under a known scaling or FIR filtering behavior for an impulse), the block output is compared to a reference result within a specified tolerance.
- **Contract checks:** The block verifies required input fields (domain tags, sampling information) and ensures that output fields are correctly updated (e.g., the sampling rate after resampling stages; impairment metadata appended when enabled).
- **Calibration-readiness checks:** For blocks intended for calibration to measurements, the code verifies that the parameterization exposes identifiable knobs (such as gain, bandwidth, noise bandwidth, nonlinearity parameters), and that “nominal vs calibrated” settings can be switched without altering code paths.

This approach is consistent with best practices in scientific software engineering: reliability requires systematic testing, modular design, and early detection of incorrect assumptions, especially when a codebase is intended to be reused and extended over time⁷⁴.

4.6.4 End-to-end regression tests (golden cases, pass/fail criteria)

Finally, the framework employs end-to-end regression tests to ensure that improvements in one component do not inadvertently disrupt previously validated behavior. The main idea is to keep a small set of *golden cases* (fixed scenario IDs, parameters, and seeds) that represent typical operating points (e.g., baseline PAM case, bandwidth-limited case, calibrated-parameter case). Each golden case defines:

- **Inputs:** full parameter snapshot, seed(s), probe points, and enabled blocks.

⁷⁴ H. Hunter-Zinck, A. F. de Siqueira, V. N. Vásquez, R. Barnes, and C. C. Martinez, "Ten simple rules on writing clean and reliable open-source scientific software," PLoS Computational Biology, vol. 17, no. 11, e1009481, 2021.

- **Expected observables:** a compact set of metrics such as BER/EVM (when Rx DSP is enabled), eye statistics, spectrum-based observables, and summary waveform statistics at agreed probe points.
- **Pass/fail criteria:** absolute tolerances for deterministic quantities, and statistically justified tolerances for stochastic quantities (e.g., metrics computed over a finite number of bits).

The logic of defining simulation objectives, selecting metrics, fixing seeds where needed, and documenting evaluation conditions reflects established guidelines for rigorous simulation studies: results should be reproducible, comparisons should be fair, and stochastic variability must be addressed explicitly rather than implicitly⁷⁵. Additionally, maintaining a documented protocol for what “valid enough” means at each development stage aligns with broader validation methodologies used in simulation disciplines: validity is not a binary label but a structured process tied to purpose, observables, and documented assumptions⁷⁶.

⁷⁵ T. P. Morris, I. R. White, and M. J. Crowther, "Using simulation studies to evaluate statistical methods," *Statistics in Medicine*, vol. 38, no. 11, pp. 2074-2102, 2019.

⁷⁶ C. Troost, R. Huber, A. R. Bell, H. van Delden, T. Filatova, Q. B. Le, M. Lippe, L. Niamir, J. G. Polhill, Z. Sun, and T. Berger, "How to keep it adequate: A protocol for ensuring validity in agent-based simulation models," *Environmental Modelling & Software*, vol. 159, 105559, 2023.

Chapter 5

Transmitter-Side Modeling: Digital, Electrical, Optical

High-speed short-reach IM/DD links are increasingly operated in regimes where transmitter non-idealities are no longer “implementation detail,” but first-order performance drivers: the achievable BER/eye opening depends on the coupled action of digital shaping, mixed-signal conversion, and electro-optic transfer, not only on symbol-rate link abstractions. Accordingly, this chapter treats the transmitter as a sampled, mixed-domain waveform generator, where pre-DAC DSP, DAC imperfections (static and dynamic), electrical front-end bandwidth/noise, and modulator nonlinearity jointly determine the waveform presented to the channel and, ultimately, the receiver equalizer’s burden. This viewpoint aligns with the short-reach IM/DD DSP literature, in which practical feasibility hinges on waveform-level effects under bandwidth/quantization constraints rather than on idealized symbol models¹. It is also consistent with the technology roadmaps for 100–200+ Gb/s-per-lane IM/DD, which explicitly emphasize tight margins and hardware-aware design².

Within the architecture defined in Chapter 4, Chapter 5 documents the transmitter-side surrogate blocks as a coherent modeling stack: (i) digital source and mapping for PAM/BPAM with explicit rate and sampling definitions; (ii) a DAC surrogate that includes remapping to physical range, quantization, and calibrated non-idealities (e.g., INL/DNL and dynamic artifacts) as parameterized, identifiable impairments; (iii) the driver/electrical front-end as the bridge between

¹ T. Wettlin, S. Calabrò, T. Rahman, J. Wei, N. Stojanovic, S. Pachnicke, “*DSP for High-Speed Short-Reach IM/DD Systems Using PAM*,” *Journal of Lightwave Technology*, vol. 38, no. 24, pp. 6771–6778, 2020.

² X. Pang, O. Ozolins, R. Lin, L. Zhang, A. Udalcovs, L. Xue, R. Schatz, U. Westergren, S. Xiao, W. Hu, G. Jacobsen, S. Popov, and J. Chen, “*200 Gbps/Lane IM/DD Technologies for Short Reach Optical Interconnects*,” *Journal of Lightwave Technology*, vol. 38, no. 2, pp. 492–503, 2020.

digital samples and modulator drive; and (iv) the MZM transfer model as the deterministic electro-optic nonlinearity that shapes intensity statistics before propagation. The BPAM case is treated as a distinct operating mode because feasibility under direct detection relies on waveform reconstruction principles and therefore places specific requirements on sampling, bandwidth, and distortion control³. Where higher-order device imperfections are not yet fully implemented (e.g., MZM imbalance/chirp extensions, richer channel physics beyond the current baseline), they are still defined mathematically and parameterized so the thesis remains technically complete and the code remains extensible under the same block contracts and calibration hooks described earlier⁴.

5.1 Chapter scope, assumptions, and links to the framework

This chapter details the transmitter-side part of the digital-surrogate chain outlined in Chapter 4, emphasizing the series of transformations that convert a discrete bitstream into a continuous-time optical waveform at the modulator output. In practical IM/DD links, especially at high per-lane symbol rates, the transmitter's performance is dominated by the combined effects of DSP decisions (mapping/normalization), DAC/driver imperfections, and electro-optic conversion physics. This underscores the need for a transmitter model that is both physically parameterized and probe-inspectable, enabling waveform inspection and metric extraction⁵.

At the modeling level, we follow the runtime conventions from Chapter 4: the transmitter runs as a series of blocks with a clear I/O contract (required fields,

³ M. Secondini and E. Forestieri, “*Direct Detection of Bipolar Pulse Amplitude Modulation*,” *Journal of Lightwave Technology*, Vol. 38, Issue 21, pp. 5981-5990, 2020.

⁴ X. Zhou, R. Urata, and H. Liu, “*Beyond 1 Tb/s Intra-Data Center Interconnect Technology: IM-DD OR Coherent?*,” *Journal of Lightwave Technology*, vol. 38, no. 2, pp. 475–484, 2020.

⁵ X. Pang, O. Ozolins, R. Lin, L. Zhang, A. Udalcovs, L. Xue, R. Schatz, U. Westergren, S. Xiao, W. Hu, G. Jacobsen, S. Popov, and J. Chen, “*200 Gbps/Lane IM/DD Technologies for Short Reach Optical Interconnects*,” *Journal of Lightwave Technology*, vol. 38, no. 2, pp. 492–503, 2020.

optional fields, and domain tags), and with consistent unit declarations at each boundary. The scope is limited to the Tx chain up to the optical waveform shown to the next stage of propagation, amplification, and receiver front-end modeling in Chapter 6. Even if a second-order impairment isn't yet enabled in the current implementation, we still define the mathematical representation and the parameter “hooks” that the framework offers for future enhancement.

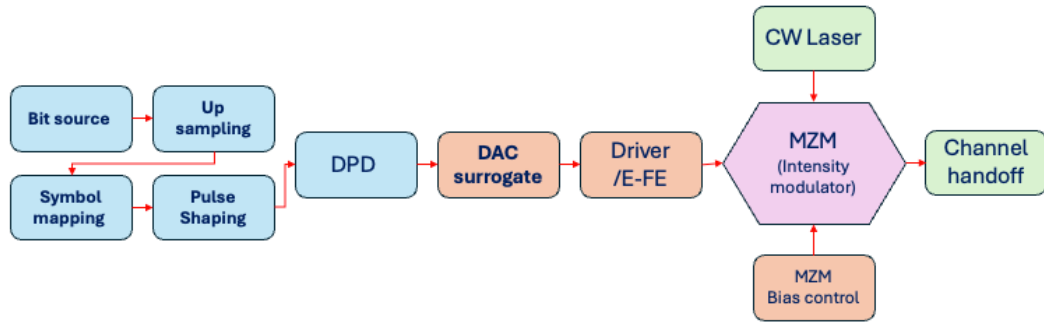


Figure 5.1 Transmitter scope and domain transitions.

5.1.1 Position in the end-to-end chain (handoff from Chapter 4 interface conventions)

In the end-to-end execution model, this chapter represents the *front segment* of the complete link.

- **Bit source and formatting** (implemented via “generate_bits” and its mode-specific helpers such as “generate_random_bits”, “generate_manual_bits”, and “generate_prbs_bits”): produces a binary sequence under either deterministic (seeded) or stochastic (shuffled) operation.
- **Symbol mapping and normalization**: maps bits to amplitude levels (PAM or bipolar-PAM), optionally applying Gray coding when necessary, and establishes the reference scaling used by subsequent analog-emulation stages.
- **Digital-to-analog conversion surrogate (DAC)**: performs range remapping to a physical voltage span, then applies quantization and selected analog-emulation effects (including static and dynamic non-idealities and noise) within a block-local, parameterized structure.

- **Driver/electrical front end:** enforces bandwidth constraints and gain, and optionally contributes electrical noise and nonlinear response.
- **Laser + electro-optic conversion (MZM + bias control):** generates the carrier field (complex envelope) and modulates it based on the MZM transfer function under a specific bias condition; the resulting optical waveform is then sent to the channel stage.

This breakdown mirrors the architecture commonly used in high-speed short-reach IM/DD systems, where the *DSP + mixed-signal Tx chain* is treated as a coupled subsystem and evaluated via waveform-level observables such as eye opening, noise statistics, and effective SNR/BER proxies⁶. It also matches the practical engineering perspective that, as baud rate increases, the “useful” performance envelope is increasingly limited by *analog bandwidth*, *DAC/driver linearity*, and *modulator biasing*, rather than by idealized symbol mapping alone⁷.

5.1.2 Domain transitions and unit conventions used in this chapter (digital/electrical/optical)

This chapter operates across three modeling domains; the purpose of explicitly declaring domain transitions is to prevent silent unit inconsistencies and to make the output of every block interpretable as a “virtual instrument probe.”

In the *digital* domain (dimensionless, discrete-time), the bitstream $b[n] \in 0,1$ and symbol sequence $a[k]$ are treated as dimensionless sequences. A waveform representation is obtained through a defined sampling framework:

- *Bit rate* R_b [bit/s] and *symbol rate* R_s [baud] determine the fundamental unit interval $T_s = 1/R_s$.

⁶ T. Wettlin, S. Calabrò, T. Rahman, J. Wei, N. Stojanovic, S. Pachnicke, “*DSP for High-Speed Short-Reach IM/DD Systems Using PAM*,” Journal of Lightwave Technology, vol. 38, no. 24, pp. 6771–6778, 2020.

⁷ X. Pang, O. Ozolins, R. Lin, L. Zhang, A. Udalcovs, L. Xue, R. Schatz, U. Westergren, S. Xiao, W. Hu, G. Jacobsen, S. Popov, and J. Chen, “*200 Gbps/Lane IM/DD Technologies for Short Reach Optical Interconnects*,” Journal of Lightwave Technology, vol. 38, no. 2, pp. 492–503, 2020

- A chosen *sampling rate* F_s [Hz] defines the sample period $T = 1/F_s$ and the *samples-per-symbol* $\text{SpS} = F_s/R_s$ (integer in the nominal configuration).

- Any upsampling, pulse shaping, or pre-emphasis steps are expressed as discrete-time operations whose outputs remain dimensionless until the DAC remapping stage defines a physical voltage range.

In the *electrical* domain (voltage-referenced, discrete-time), at the DAC output, the waveform is represented as an electrical signal $v[n]$ in volts [V]. The conversion from digital to electrical is made explicit by introducing a *full-scale range* (or equivalent voltage span) and mapping the normalized symbol waveform into that range. This is the point where mixed-signal impairments are defined in physically meaningful terms.

- Quantization is parameterized by an effective resolution N (bits) and a step size Δ consistent with the configured voltage range.

- Driver and electrical front-end operations preserve volt units, applying gain, filtering (band-limitation), and optional nonlinear or noise terms that are explicitly stated as additive (e.g., $v[n] + \eta[n]$) or multiplicative (e.g., $gv[n]$).

In the *optical* domain (field and power, baseband-equivalent), the optical carrier is represented through a complex field envelope $E(t)$ and/or optical power $P(t)$, depending on the modeling granularity of the stage:

- A CW laser stage provides $E_{\text{in}}(t)$ an optional phase evolution (e.g., drift/random walk) while remaining in a baseband-equivalent representation (i.e., without requiring optical-frequency sampling).

- The MZM implements an intensity-modulation transfer function driven by the electrical waveform and a bias condition, producing $E_{\text{out}}(t)$, and the corresponding instantaneous power $P(t) = |E_{\text{out}}(t)|^2$. The bias point is treated as a first-class configuration element because practical MZMs exhibit operating-point drift; consequently, bias control strategies (e.g., based on monitoring OMA and average optical power) are required to maintain stable modulation conditions across long acquisitions and

parameter sweeps⁸. Moreover, the *target* operating point is modulation-dependent: while conventional unipolar IM/DD PAM is typically biased in the quadrature point to preserve quasi-linear intensity mapping, BPAM with direct detection intentionally operates the modulator at (or near) the **null** to enable the bipolar-to-intensity mapping exploited by the receiver⁹.

Accordingly, the surrogate treats the MZM-bias source/controller as an explicit block that can provide (i) a deterministic set-point (quadrature vs. null) and (ii) statistical perturbation models that emulate practical bias instabilities, such as slow drift (e.g., random-walk / low-frequency wander) and slight bias jitter around the set-point. These bias perturbations can then be swept or randomized to measure effects on eye opening, BER/EVM, and level statistics, and importantly, to create labeled datasets for optimization or machine learning training where bias fluctuations are included as part of the operating conditions manifold.

These conventions ensure that (i) the meaning of each intermediate waveform is clear (dimensionless vs volts vs field/power), and (ii) subsequent “virtual instrumentation” (time-domain plots, eye diagrams, spectra, and metric extraction) can be applied consistently at any probe point without guessing hidden scale factors or implicit units.

5.2 Digital transmit processing and symbol formation

This section specifies the *digital transmitter front end* that generates the discrete-time sequence used by downstream symbol-formation and waveform-synthesis blocks. In the proposed surrogate, bitstream generation is treated as a *standalone source module*: it outputs a binary sequence with a defined length and reproducibility policy, independent of the chosen modulation format, pulse shaping, or physical-layer impairments. This separation is deliberate; future additions such

⁸ M. H. Kim, B. M. Yu, and W. Y. Choi, “A Mach–Zehnder Modulator Bias Controller Based on OMA and Average Power Monitoring,” *IEEE Photonics Technology Letters*, vol. 29, no. 23, pp. 2043–2046, 2017.

⁹ M. Secondini, E. Forestieri, A. Fresi, L. Poti, A. Catuogno, and F. Cavaliere, “Bipolar pulse amplitude modulation with direct detection,” in *45th European Conference on Optical Communication (ECOC 2019)*, paper We4.61, 2019.

as framing, FEC, and alternative mappers can be inserted without altering the semantics of the upstream bit source, while still preserving the global rate configuration (e.g., " R_b , R_s , F_s , and samples-per-symbol definitions) enforced by the Chapter 4 runtime conventions.

5.2.1 Bit sources and reproducibility (import / random / IEEE PRBS, seed policy)

The bit-source stage is implemented by the wrapper “generate_bits”, which writes the generated sequence into the runtime container and initializes the digital-domain state required by subsequent blocks. Concretely, the module produces a binary column vector:

$$\mathbf{b} \in \{0,1\}^{N_b},$$

stores it as “signal.bits”, and mirrors it into “signal.current” as a “double” vector for downstream numerical processing. The output domain is explicitly labeled as **digital** (via “signal.domain = 'digital'”), and the block is designed to be *downstream-agnostic*: it does not assume PAM vs. QAM vs. PSK, unipolar vs. bipolar signaling, or any particular framing structure.

The selection of the bit-source mode is controlled centrally through the configuration fields:

- “parameters.bitgen.mode” (e.g., 'manual', 'random', 'prbs', 'import')
- “parameters.bitgen.nBits” (target length (N_b))
- “parameters.bitgen.useSeed, parameters.bitgen.seed”

(reproducibility policy)

- “parameters.bitgen.prbsOrder, parameters.bitgen.prbsShift” (PRBS configuration)
- “parameters.bitgen.manualBits” (explicit user-provided bits)

(i) Manual bitstream (deterministic explicit input).

In “manual” mode, “generate_manual_bits” validates that the provided vector contains only 0,1, forces a column orientation, and then constructs a length- N_b stream by *periodic repetition* if needed. This mode is intended for deterministic micro-tests (e.g., debugging mapping edge-cases, validating framing hooks, or

verifying waveform evolution under a known binary pattern), where reproducibility must be absolute and independent of random-number state.

(ii) Random bitstream (stochastic source with explicit seed policy).

In random mode, “generate_random_bits” produces an independent and identically distributed (i.i.d.) binary stream consistent with a *Bernoulli(0.5)* source (implemented as `bits = randi([0 1], nBits, 1)`), suitable for Monte-Carlo BER/SER studies. Reproducibility is enforced through an explicit two-state RNG policy using MATLAB’s RNG infrastructure (Mersenne Twister). If “parameters.bitgen.useSeed = true”, the generator is initialized with `rng(seed,'twister')`, so the same (*seed, configuration, code version*) deterministically reproduces the identical bitstream (scenario replay); if `parameters.bitgen.useSeed = false`, `rng('shuffle')` is invoked (time-dependent seeding), yielding non-repeatable sequences intended for exploratory runs and broad statistical sampling. Crucially, the selected mode and (when applicable) the seed are treated as part of the run configuration (and recorded in the run metadata), so that stochasticity remains controllable, auditable, and exactly re-executable—an established requirement in reproducible computational practice when numerical results depend on randomized components¹⁰.

(iii) PRBS bitstream (maximal-length LFSR families with order selection).

In PRBS mode, the transmit stimulus is a *pseudorandom binary sequence* generated by an m -stage linear feedback shift register (LFSR) driven by a primitive feedback polynomial. At each clock cycle, a feedback bit is computed as the XOR of selected “tap” stages defined by the polynomial; the register is shifted, and one stage is taken as the output bit. When the feedback polynomial is primitive, and the initial register state is non-zero, the LFSR produces a maximal-length sequence with period $2^m - 1$. The all-zero state is avoided because it is a lock-up state (it would output zeros forever). In practice, PRBS families (e.g.,

¹⁰ G. K. Sandve, A. Nekrutenko, J. Taylor, and E. Hovig, “Ten Simple Rules for Reproducible Computational Research,” *PLoS Computational Biology*, vol. 9, no. 10, e1003285, 2013

PRBS7/9/11/15/23/31) are widely used as compliance-style stress patterns because they approximate broadband excitation while remaining exactly repeatable¹¹.

In the current implementation, “generate_prbs_bits” realizes this structure using MATLAB’s “comm.PNSequence”, configured by (i) a primitive polynomial selected by “getPrimitivePolynomial(prbsOrder)” and (ii) non-zero initial conditions (all ones) to fix the PRBS phase and prevent lock-up. The generator is requested to output a frame of length “nBits” (so the sequence length is controlled explicitly by the simulation configuration). A deterministic phase/offset control is also supported via “parameters.bitgen.prbsShift”, implemented as a repeatable cyclic offset by discarding the first shift bits and then using the subsequent bits; this is useful when aligning the PRBS phase with external measurement conventions or when avoiding early transient effects during staged debugging.

In high-speed optical interconnect and Ethernet-oriented contexts, the term “IEEE PRBS” is commonly used to denote PRBS patterns specified in IEEE Ethernet PHY test methodologies for transmitter/receiver stress and compliance testing (including long patterns such as PRBS31 and, on PAM4 lanes, defined variants such as PRBS13Q)¹². For this reason, the PRBS option is not merely a convenience feature: it is the mechanism by which simulation stimuli can be aligned with standardized test practices, enabling controlled stress of pattern-dependent impairments and facilitating apples-to-apples comparison between simulated and laboratory results.

(iv) Import mode (explicit placeholder for file-driven sources).

An “import” mode is declared in the wrapper interface, but it is intentionally left as a hard error in the current implementation (“error('Bit import mode is not implemented yet.').”). This is consistent with the plug-and-play architecture: the interface contract is already fixed, and a future importer can be added without

¹¹ ITU-T, “*Digital test patterns for performance measurements on digital transmission equipment*,” ITU-T Recommendation O.150, Oct. 1996.

¹² IEEE, “*IEEE Standard for Ethernet-Amendment 10: Media Access Control Parameters, Physical Layers, and Management Parameters for 200 Gb/s and 400 Gb/s Operation*,” IEEE Std 802.3bs-2017, 2017.

changing downstream assumptions (the only required output remains $\mathbf{b} \in \{0,1\}^{N_b}$ with an explicit digital-domain tag).

Output contract and downstream compatibility. Across all modes, the bit-source module guarantees:

1. Binary validity: values are only 0 or 1.
2. Length control: output length equals “parameters.bitgen.nBits”.
3. Deterministic reproducibility when requested: controlled exclusively via “useSeed/seed” (random mode) or deterministic generation (manual/PRBS).
4. Domain clarity: the signal is explicitly flagged as digital at the output, enabling strict downstream domain transitions (digital-to-electrical-to-optical) to be enforced later in the chain.

5.2.2 Mapping for PAM and BiPAM/BPAM (including polarity handling)

In the proposed framework, symbol mapping is seen as a standalone DSP interface between the bit source and all downstream transmitter blocks: it transforms a binary stream into a dimensionless discrete-time symbol sequence with a clearly defined alphabet, bit ordering, and labeling rule (binary vs. Gray). This separation is intentional: it allows the same upstream bit sources and framing hooks to feed different modulation families, while the downstream chain (pulse shaping, DAC remap, electrical/optical conversion) works with a standardized symbol representation.

Implemented modulation dispatch and “plug-in” structure

The entry-point function selects the modulation family from parameters and routes the request to the corresponding mapper. In the current codebase, PAM mapping is implemented. In contrast, QAM and PSK are present as extension stubs that intentionally raise an error (to make incompleteness explicit while preserving the modular interface).

Bit packing, symbol indexing, and padding policy

Let the modulation order be M . For PAM mapping, the mapper requires a bit-to-symbol arity

$$k = \log_2(M),$$

and therefore enforces that M be a power of two (except for a dedicated future branch for $M = 6$, see below). The input bitstream $b \in 0,1^{N_b}$ is then packed into a matrix of N_s symbols and k bits per symbol by reshaping and using left-MSB interpretation:

$$i_n = \text{bi2de}(b_n, \text{left-msb}), \quad i_n \in 0, \dots, M - 1.$$

A practical detail implemented in the mapper function is that if N_b is not an integer multiple of k , the function pads the stream with a small number of random bits so that reshaping is well-defined; the padded bits do not change the mapper's semantics but prevent silent dimensional inconsistencies during large batch runs.

Binary vs. Gray labeling (nearest-neighbor error mitigation)

After forming the natural binary symbol index i_n , the mapping stage optionally applies *Gray labeling*. In the implementation, Gray indices are generated by the standard XOR rule

$$g_n = i_n \oplus \left\lfloor \frac{i_n}{2} \right\rfloor,$$

and then converted to MATLAB's 1-based indexing by adding 1 before alphabet lookup. This is the canonical Gray-code construction used to ensure that *adjacent amplitude levels differ by one bit*, thereby reducing the average bit error impact of the most probable nearest-neighbor symbol errors in noise-limited regimes¹³.

PAM alphabet generation and polarity handling (PAM vs. BiPAM/BPAM)

The same PAM mapper supports both conventional nonnegative (unipolar) PAM and symmetric-around-zero (bipolar) PAM by passing a *polarity flag*. The alphabet is generated by a dedicated subfunction and then normalized to a consistent full-scale reference:

¹³ J. G. Proakis and M. Salehi, "Digital Communications," McGraw-Hill Education, 5th ed., 2008.

$$a \leftarrow \frac{a}{\max(|a|)}.$$

This yields a dimensionless symbol sequence s_n that is bounded in $[-1,1]$ for bipolar and $[0,1]$ for unipolar operation, which is advantageous because later blocks (DAC remap, driver gain, modulator transfer function) can impose physical scaling without redefining the abstract DSP alphabet.

Concretely, the alphabet construction implemented in the standard PAM alphabet follows the standard equally spaced level sets: for bipolar PAM (restricted to even M in the current function),

$$a_m = 2m - (M - 1), \quad m = 0, \dots, M - 1,$$

which is then normalized by $(M - 1)$. For unipolar PAM,

$$a_m = m, \quad m = 0, \dots, M - 1,$$

again normalized by $(M-1)$. This polarity switch is exactly where the framework distinguishes “ordinary” PAM (often realized downstream as intensity levels) from BPAM, where the DSP symbols are intentionally zero-centered so that their sign is preserved through the transmitter’s digital/electrical representation.

From a system perspective, *BPAM differs from standard IM/DD PAM not by a special mapper*, but by (i) choosing a bipolar alphabet at the DSP layer and (ii) making sure that subsequent physical-domain blocks (DAC remap, biasing strategy, optical modulation/detection assumptions) maintain and utilize this bipolar structure. This design choice aligns with BPAM direct-detection formulations where the transmitted waveform is regarded as bipolar at the electrical/DSP abstraction level, and the link design (including modulator operating point strategy) is engineered accordingly¹⁴.

Explicit implementation boundary: PAM6 placeholder

The mapper includes a conditional branch for $M=6$, but that function is currently a non-implemented placeholder that raises an error. This is consistent with the overall policy of keeping the interface stable (so PAM6 can be plugged in later) while preventing accidental use without a validated alphabet/labeling rule.

¹⁴ M. Secondini and E. Forestieri, “Direct Detection of Bipolar Pulse Amplitude Modulation,” *Journal of Lightwave Technology*, vol. 38, no. 21, pp. 5981–5990, 2020.

5.2.3 Optional FEC placeholder and framing hooks (pass-through by design)

In the current transmitter framework, forward error correction (FEC) is considered an *optional* pre-mapping coding stage that operates on the binary payload and generates a coded bitstream for subsequent symbol formation. Even when the encoder is turned off (current setup: pass-through), the software provides explicit “hooks” for (i) frame construction (payload length, parity overhead, alignment markers), (ii) deterministic interleaving/de-interleaving, and (iii) configuration logging (code family, code rate, block length, and related seeds). This is intentional: once an IM/DD link operates near its sensitivity limits, *post-FEC* performance, not just uncoded BER, becomes the key metric, and the simulation must be able to replicate the exact framing/coding context used for any reported BER/SNR analysis.

From a communications standpoint, FEC introduces structured redundancy by mapping a length- k information block into a length- n codeword (rate $R = k/n$; overhead

$OH = (n - k)/k$). Decoding exploits this redundancy to correct errors produced by the physical channel and receiver processing. Modern optical systems typically distinguish *hard-decision* (HD-FEC) from *soft-decision* (SD-FEC): HD-FEC decoders operate on quantized symbol decisions (low latency/complexity), whereas SD-FEC decoders exploit reliability information (e.g., Log-Likelihood Ratios, LLRs) from the demapper for higher coding gain at the expense of computational cost and power¹⁵. In a modular surrogate, this distinction is structural: SD-FEC requires a demapper interface that exports soft metrics, whereas HD-FEC only requires a binary decision stream. Therefore, the FEC block boundary must be explicitly specified in the block-I/O contract, even before the encoder is implemented.

For short-reach PAM4-based interconnects, including Ethernet-class links, the common practice is HD-FEC based on Reed–Solomon block codes standardized at

¹⁵ G. Tzimpragos, C. Kachris, I. B. Djordjevic, M. Cvijetic, D. Soudris, and I. Tomkos, “A Survey on FEC Codes for 100G and Beyond Optical Networks,” IEEE Communications Surveys & Tutorials, vol. 18, no. 1, pp. 209–221, 2016.

the PHY/PCS level, such as the widely used KP4 RS-FEC¹⁶. This approach (coding before mapping and decoding after detection) aligns with the pass-through design used here: the bit source and subsequent symbol mapping do not need redesigning when FEC is enabled; only the inserted encoder/decoder blocks and framing metadata become active. The specific modular placement also supports future “stress-pattern” experiments, in which pattern-dependent impairments lead to bursty error statistics. In this scenario, interleaver depth and codeword boundaries can be as influential as the uncoded BER itself. Therefore, the framework must retain these degrees of freedom rather than reducing everything to a single uncoded BER figure.

Beyond Ethernet-class HD-FEC, higher-gain FEC families used in optical transport include iterative product-like constructions such as *staircase codes*, which combine substantial net coding gain with implementable HD decoding complexity and are widely used as a practical compromise between classical RS codes and SD-FEC¹⁷. When the thesis progresses to receiver-side DSP and end-to-end optimization, these code families define clear future integration targets: (i) KP4-style RS-FEC as the baseline “standards-aligned” option for PAM-style short-reach; and (ii) staircase-class codes when coding gain is prioritized, and the system budget permits iterative decoding.

Implementation note (current state). The current codebase keeps the FEC block as a pass-through, but the transmitter already carries the correct *conceptual insertion point* and *framing interface* to activate it later without refactoring downstream blocks. In practice, this means: (1) the bitstream is generated independently of downstream choices; (2) a framing layer can be enabled to define payload/codeword segmentation; and (3) the run configuration can be extended to log FEC identity and parameters so that post-FEC metrics can be reproduced exactly and used as labels/targets in ML-driven optimization.

¹⁶ IEEE, “*IEEE Standard for Ethernet*,” IEEE Std 802.3 (latest consolidated edition; see RS-FEC/KP4 clauses for PAM4 Ethernet PHYs), IEEE, 2022.

¹⁷ B. P. Smith, A. Farhood, A. Hunt, F. R. Kschischang, and J. Lodge, “*Staircase Codes: FEC for 100 Gb/s OTN*,” *Journal of Lightwave Technology*, vol. 30, no. 1, pp. 110–117, 2012.

5.2.4 Pulse shaping / upsampling strategy and reference rate definitions

After symbol mapping, the transmitter creates a discrete-time waveform that can be (i) examined in the digital domain (such as time plots and eye diagrams) and (ii) sent downstream to the DAC surrogate under a clearly defined time base. In the current codebase, this stage is implemented as two separate, modular blocks: zero-insertion upsampling followed by Nyquist pulse shaping. This setup ensures that the waveform bandwidth, ISI behavior, and sampling granularity are determined by configuration, rather than being implicit in downstream analog models.

Reference rates and derived sampling. Let the bit rate be R_b [bit/s] and the modulation carry $k = \log_2(M)$ bits/symbol (e.g., PAM-(M)); then the symbol rate is

$$R_s = \frac{R_b}{k} \quad [\text{sym/s}].$$

The internal discrete-time sampling rate is defined by the samples-per-symbol parameter $\text{SpS} \in \mathbb{Z}^+$, yielding

$$F_s = R_s \cdot \text{SpS} \quad [\text{Hz}], \quad T_s = \frac{1}{F_s}.$$

This choice makes the sampling grid explicit and stable for subsequent visualization (time plots, spectra) and for eye-diagram folding based on the unit interval (UI). In the implementation, F_s is set immediately after upsampling as $F_s = R_s \times \text{SpS}$, so the time manager can rebuild/validate the global time base consistently once the sample count changes.

Upsampling (zero insertion). The upsampling block does not interpolate by itself; instead, it inserts $\text{SpS} - 1$ zeros between adjacent symbols:

$$X_{\uparrow}[n] = \begin{cases} a[m], & n = m, \text{SpS}, \\ 0, & \text{otherwise} \end{cases}$$

where $a[m]$ is the mapped symbol sequence, and $x_{\uparrow}[n]$ is the upsampled sequence. This is the canonical multirate front-end used before an interpolation (pulse-shaping) filter: the spectral images introduced by zero insertion are subsequently removed by the shaping filter, while the symbol timing grid is preserved exactly. In the code, this behavior is implemented directly as “place symbols at indices 1:SpS:end and zeros elsewhere”, with robust validation of SpS and explicit creation of “signal.upsampled.”

Pulse shaping (RC/RRC families). Pulse shaping is applied to $x_{\uparrow}[n]$ using standard Nyquist families to control occupied bandwidth and ISI. The framework supports raised-cosine (RC) and root-raised-cosine (RRC) pulse shaping parameterized by roll-off ($\alpha \in [0,1]$) and span (in symbols) N_{span} . Conceptually, the shaped waveform is

$$x_{\text{sh}}[n] = (x_{\uparrow} * h)[n],$$

where $h[n]$ is the FIR approximation of the selected RC/RRC pulse (Nyquist criterion; bandwidth set by α ; filter length set by N_{span} and (SpS)¹⁸.

In the implementation, filter taps are generated via an RC/RRC design call (RC: “normal”, RRC: “sqrt”), and the waveform is filtered by convolution with length preservation, so the output signal vector remains aligned to the existing sample index/time grid while the effective group delay is handled implicitly at the waveform level. The filter taps are stored in metadata to preserve the exact reproducibility of the shaping stage.

Internal sampling targets and downstream decoupling. A key architectural point is that (R_s, SpS, F_s) define the digital waveform formation grid, but they do not necessarily equal the DAC’s internal sampling target. The DAC surrogate includes an explicit sampling-rate-matching stage (resampling to the DAC rate) before applying pre-emphasis, static transfer effects (quantization + INL/DNL), and continuous-time emulation elements (ZOH, analog filtering, noise). Therefore, waveform formation (upsampling + pulse shaping) is treated as a modular, inspectable, and optimizable stage whose output can be replayed exactly, while DAC-rate enforcement and analog bandwidth constraints remain localized inside the DAC block.

Implications for optimization/ML hooks (forward link). Because roll-off, span, and other parameters directly influence spectral occupancy, intersymbol interference, and discretization granularity, these are natural “knobs” for later optimization loops (such as balancing bandwidth, eye opening, and sensitivity to nonlinearities/noise). In this setup, these knobs are already visible in the parameter tree, so they can be logged as part of scenario identity and provenance. This enables

¹⁸ J. G. Proakis and M. Salehi, “*Digital Communications*,” McGraw-Hill, 5th ed., 2008.

reproducible sweeps and ML-ready dataset creation when the optimization layer is activated.

5.3 DAC surrogate block (digital front-end + converter core)

The DAC module functions as a mixed-signal boundary block that transforms a discrete-time, numerically defined transmit waveform into a continuous-time electrical drive signal suitable for injection into the downstream analog chain (driver and MZM). Instead of approximating the DAC as an ideal “sampling + quantization” operator, it is modeled as a structured sequence of digitally defined conditioning operations and converter-core effects, with their order chosen to reflect how impairments accumulate in real hardware. This intentional positioning is essential because many transmitter phenomena that dominate short-reach IM/DD performance, such as spectral shaping, intersymbol memory, amplitude compression, and pattern-dependent distortion, are either generated or significantly amplified at the DAC boundary, making it essential for these effects to be localized within a block whose inputs, outputs, and parameters remain physically interpretable.

The DAC surrogate is structured as a pipeline with clear injection points for (i) internal digital conditioning (such as sampling-rate matching and optional pre-emphasis), (ii) hardware-aware remapping into the converter’s voltage range, (iii) quantization and code mapping with a specified clipping policy, (iv) static transfer deviations (INL/DNL), and (v) reconstruction behavior (ZOH and analog bandwidth limitation), followed by physically parameterized noise and timing perturbations where relevant. The result is an electrical waveform whose statistics and spectrum align with both the configured operating point and the modeled non-idealities, while maintaining traceability. Each effect can be traced to a specific mechanism and parameter group, and the block can optionally output structured metadata for future calibration and machine learning/optimization workflows without affecting the primary signal state.

5.3.1 Role of the DAC in the surrogate and why device-level modeling matters

In an IM-DD transmitter chain, the DAC is not merely a passive “format converter”; it is the primary interface at which discrete-time DSP outputs are converted into the continuous-time electrical stimulus that ultimately drives the electro-optic front end. The DAC jointly influences (i) the achievable amplitude granularity (effective resolution/ENOB), (ii) the available time discretization for the analog reconstruction process (sampling rate and hold behavior), and (iii) the deterministic distortion mechanisms shaping the electrical waveform before reaching the driver and modulator. Therefore, assuming the DAC as ideal can lead to transmitter waveforms that are numerically “valid” in the DSP domain but physically impossible once finite resolution, code-dependent nonlinearity, and reconstruction dynamics are considered.

A key motivation for device-level DAC modeling is that modern short-reach optical interconnect transmitters operate in regimes where the performance margin is often dictated by transmitter hardware constraints, such as bandwidth and nonlinear distortion, rather than channel noise. Experimental and comparative IM-DD studies have consistently shown that the performance achievable with multi-level intensity-modulated formats depends heavily on practical transmitter-side waveform generation and conditioning, not just on the abstract modulation alphabet¹⁹. In this scenario, the DAC becomes a primary factor in determining effective eye opening, waveform symmetry, and ultimately BER, especially at high symbol rates where the electrical chain is bandwidth-limited, and the reconstruction process becomes significant.

From a modeling standpoint, DAC impairments are usefully separated into (a) *static transfer effects* and (b) *dynamic reconstruction effects*. Static effects include quantization, clipping, and code-dependent non-idealities such as INL/DNL and mismatch, which translate directly into deterministic amplitude errors and constellation/level distortion. Analytic treatments of DAC mismatch and its induced constellation statistics make explicit that non-ideal DAC behavior can be

¹⁹ K. Zhong, X. Zhou, T. Gui, L. Tao, Y. Gao, W. Chen, J. Man, L. Zeng, A. P. T. Lau, and C. Lu, “Experimental study of PAM-4, CAP-16, and DMT for 112-Gb/s short-reach optical transmission systems,” *Optics Express*, vol. 23, no. 2, pp. 1176–1185, 2015.

mapped to *measurable degradation in metrics such as EVM* (and, by extension, SNR-like observables), providing a principled justification for representing these effects inside the surrogate rather than absorbing them into a generic “implementation penalty²⁰.”

Dynamic converter-core effects are modeled at the point where the discrete-time sequence is converted into a continuous-time electrical drive waveform. In particular, the DAC’s sample-and-hold operation is represented as a zero-order hold (ZOH) with period $T_s = 1/F_s$, i.e., a piecewise-constant hold that is equivalent to convolution with a rectangular pulse in time and therefore imposes the well-known $\text{sinc}(\sin x/x)$ amplitude response and generates spectral images around integer multiples of F_s . The subsequent output-path bandwidth limitation and reconstruction/anti-imaging filtering (whether explicit or implicit in the DAC output buffer) further shapes amplitude and phase; in symbol-rate terms, the combined ZOH+analog output response defines the effective transmit pulse shape delivered to the driver/modulator chain and thus introduces deterministic memory (ISI) that must be co-designed with digital pre-emphasis. Neglecting these stages in a surrogate systematically biases both eye-domain observables (sampling-phase sensitivity, eye closure/asymmetry) and spectrum-domain observables (roll-off, image suppression, and in-band noise estimates)²¹. In addition, glitch transients are treated as switching-time disturbances superimposed on the ZOH waveform during code transitions; while not part of the linear ZOH response, they inject broadband energy and can create pattern-dependent spurs and incremental eye degradation²².

Device-level modeling is also crucial because many performance-prediction and mitigation workflows implicitly assume that converter limitations are correctly represented with the proper spectral structure and injection point. Frequency-dependent variations in DAC ENOB have been shown to significantly affect the accuracy of predicted high-speed eye diagrams, implying that a surrogate must

²⁰ K. Zhong, X. Zhou, T. Gui, L. Tao, Y. Gao, W. Chen, J. Man, L. Zeng, A. P. T. Lau, and C. Lu, “Experimental study of PAM-4, CAP-16, and DMT for 112-Gb/s short-reach optical transmission systems,” *Optics Express*, vol. 23, no. 2, pp. 1176–1185, 2015.

²¹ W. Kester, ed., “*The Data Conversion Handbook*,” Elsevier/Newnes, 2005, ISBN: 0750678410

²² *Ibidem*.

capture frequency-dependent converter limitations rather than treating the DAC as an ideal quantizer with a fixed ENOB²³. In other words, an impairment model that is too simplistic might still match a narrow range of waveforms, but it will falter when the transmitter operates at new points, such as different symbol rates, scaling, clipping margins, or shaping, especially during optimization.

Finally, the DAC block is pivotal for the thesis’s longer-term objective of *AI-assisted optimization*. Any optimizer trained on surrogate outputs will learn the statistical structure of impairments present in the training data; if the DAC is idealized, the optimizer will converge toward solutions that exploit degrees of freedom that do not survive hardware realization²⁴. Conversely, when DAC behavior is modeled with physically meaningful parameters (resolution, transfer nonlinearity statistics, reconstruction behavior, and noise), the optimization problem becomes properly constrained by implementability, and the metadata produced by the DAC block becomes a meaningful descriptor of the “hardware state” associated with each experiment²⁵.

5.3.2 Sampling-rate matching and digital conditioning inside the DAC module

In the proposed surrogate, the DAC block is not regarded as a “memoryless quantizer” but as a converter subsystem that includes a digital front end (DFE) and the converter-core interface. As a result, the initial operations within the DAC module create a time-consistent sampled representation (sampling-rate matching) and apply optional linear digital conditioning (pre-emphasis) before the converter-core stages (code mapping, INL/DNL, ZOH/reconstruction, and noise injection).

²³ A. Safarnejadian, W. Shi, L. A. Rusch, and M. Zeng, “Frequency-Dependent Variations in DAC ENOB: Predicting Accurate High-Speed Eye-Diagrams,” *IEEE Photonics Technology Letters*, vol. 37, no. 22, pp. 1297–1300, 2025.

²⁴ M. Rasch et al., “Hardware-aware training for large-scale and diverse deep learning inference workloads using in-memory computing-based accelerators,” *Nature Communications*, vol. 14, Art. 967, 2023.

²⁵ M. Nazarathy and I. Tomkos, “All-Analytic Statistical Modeling of Constellations in (Optical) Transmission Systems Driven by High-Speed Electronic Digital to Analog Converters Part I: DAC Mismatch Statistics, Metrics, Symmetries, Error Vector Magnitude,” *Photonics*, vol. 11, no. 8, Art. 747, 2024.

This placement is deliberate: it aligns the simulation with how practical high-speed transmitters position linear conditioning near the conversion interface, where it can be parameterized and calibrated based on electrical measurements of the DAC output path.

Sampling-rate matching inside the DAC boundary

Let $x[n]$ denote the discrete-time waveform delivered by the upstream Tx-DSP, sampled at $F_{s,\text{in}}$. The DAC subsystem is characterized by a target sampling rate $F_{s,\text{DAC}}$ (typically chosen relative to the symbol rate R_s and an intended internal oversampling factor), because subsequent converter-core effects are defined in *physical time*. The DAC module, therefore, performs an explicit *sampling-rate conversion* to obtain $x_{\text{DAC}}[n]$ at $F_{s,\text{DAC}}$, using standard multirate principles (interpolation/decimation with appropriate anti-imaging and anti-alias filtering)²⁶.

Architecturally, performing this step within the DAC block cleanly separates concerns: upstream blocks can remain “format-agnostic” (producing a correctly normalized discrete-time sequence at their own sampling rate), while the DAC block enforces the converter’s timebase and serves as the single point where time resolution is reconciled with converter and reconstruction modeling. This avoids hidden dependencies in which the effective timebase would otherwise be shaped by whichever upstream block performed the last resampling.

Pre-emphasis as DAC-front-end conditioning (not “channel EQ”)

After sampling-rate matching, the DAC module optionally applies digital pre-emphasis, modeled as a causal FIR (or IIR) shaping filter

$$x_{\text{pre}}[n] = \sum_{k=0}^{K-1} h[k] x_{\text{DAC}}[n - k]$$

whose role is to pre-compensate the *electrical* transfer characteristics seen immediately after conversion (DAC output network, output buffer, packaging/interconnect, and the immediate driver input). In other words, pre-

²⁶ A. V. Oppenheim and R. W. Schaffer, “*Discrete-Time Signal Processing*,” 3rd ed., Pearson, 2010.

emphasis targets linear bandwidth limitations and passband tilt in the *electrical transmit chain up to (and including) the modulator drive interface*, not fiber/channel propagation. This is consistent with high-speed IM/DD practice where DSP-based pre-equalization is used to recover eye opening under combined limitations of DAC resolution and transmitter bandwidth, enabling realistic eye-diagram prediction and BER evaluation for PAM signaling²⁷.

This distinction matters because the DAC’s ideal ZOH already introduces a known spectral envelope (sinc droop) in continuous time, and the subsequent analog reconstruction path adds additional roll-off and group-delay distortion. Pre-emphasis is therefore treated as a *front-end calibration/conditioning knob* whose objective is to linearize the effective electrical frequency response presented to the driver/modulator, rather than to compensate for the optical channel.

The boundary between pre-emphasis and digital predistortion (DPD)

While pre-emphasis is fundamentally a linear conditioning operation, *digital predistortion (DPD)* is a broader transmitter-side concept aimed at compensating *nonlinear* and potentially *memory-bearing* distortions (e.g., modulator transfer-function nonlinearity, finite extinction-ratio effects, or nonlinear electrical compression). DPD is commonly implemented using nonlinear predistortion (e.g., polynomial/Volterra or block-oriented Wiener–Hammerstein models) or LUT-based mappings that approximate an inverse system:

$$x_{\text{DPD}}[n] = \mathcal{F}(x[n - k]),$$

where $\mathcal{F}(\cdot)$ may include both linear and nonlinear terms. In bandwidth-limited PAM transmitters, nonlinear predistortion is used specifically to mitigate transmitter-induced waveform distortion while preserving a repeatable stimulus, allowing controlled investigation of pattern-dependent impairments and eye

²⁷ Z. Zhou, T. Odedeyi, B. Kelly, J. O’Carroll, R. Phelan, I. Darwazeh, and Z. Liu, “Impact of Analog and Digital Pre-emphasis on the Signal-to-Noise Ratio of Bandwidth-limited Optical Transceivers,” IEEE Photonics Journal, 2020.

closure²⁸. Similarly, modulator-oriented DPD frameworks explicitly target modulator-induced nonlinearities (and related effective extinction-ratio constraints), which are not reducible to linear pre-emphasis alone²⁹.

There is therefore a controlled overlap: a DPD model typically contains a linear component (equivalent to a pre-emphasis term), but its defining purpose is compensation beyond linear spectral shaping. In the surrogate design, we keep this boundary explicit by (i) treating *pre-emphasis as a DAC-front-end operation* associated with electrical reconstruction, and (ii) retaining a distinct *DPD hook* (upstream in the Tx-DSP chain) for nonlinear or learned inverse modeling. This separation prevents misinterpretation of the pre-emphasis filter as a channel equalizer and supports clearer calibration workflows.

Relevance to ML/optimization workflows

Because sampling-rate conversion and pre-emphasis are parameterized at the DAC boundary, they become **stable, identifiable degrees of freedom** for later optimization. In a learning-driven workflow, these operations can be tuned from measured electrical responses (calibration-driven) or optimized against system-level objectives (e.g., eye opening, BER proxy metrics, or robustness under DAC non-idealities). Notably, placing these steps within the DAC module ensures that any ML-based optimizer “sees” consistent timebase and reconstruction behavior, rather than learning spurious compensations caused by timebase inconsistencies across blocks.

²⁸ T. Sasai, M. Nakamura, E. Yamazaki, A. Matsushita, S. Okamoto, K. Horikoshi, and Y. Kisaka, “Wiener-Hammerstein model and its learning for nonlinear digital predistortion of optical transmitters,” *Optics Express*, vol. 28, no. 21, pp. 30952–30963, 2020.

²⁹ A. Napoli, M. M. Mezghanni, S. Calabrò, R. Palmer, G. Saathoff, and B. Spinnler, “Digital Predistortion Techniques for Finite Extinction Ratio IQ Mach–Zehnder Modulators,” *Journal of Lightwave Technology*, vol. 35, no. 19, pp. 4289–4296, 2017.

5.3.3 Electrical-range mapping and polarity handling (bipolar/unipolar remap)

A DAC is a bounded electrical source: regardless of how symbols are represented upstream, the converter output is constrained to a finite full-scale range, and any physically meaningful surrogate must make this constraint explicit before applying converter-core effects (quantization, INL/DNL, ZOH, etc.). A convenient and general representation is to treat the DAC input as a normalized discrete-time sequence $x[n]$ (real-valued, produced by the TX-DSP), and to model the *electrical-range mapping* as an affine transformation into the DAC's admissible output interval $[V_{min}, V_{max}]$ owed by an explicit clipping rule. This “map-then-perturb” viewpoint is consistent with continuous-time DAC waveform models, in which the bounded range is the starting point, and non-idealities are subsequently superimposed³⁰.

Affine remap into the DAC range (with headroom and clipping)

Let $\Delta V = V_{max} - V_{min}$ and $V_{cm} = (V_{max} + V_{min})/2$. Define a design headroom factor of $0 < \rho \leq 1$ (e.g., $\rho = 0.9$ to keep a 10% margin from the rails) and a peak normalization ($A_{max} > 0$ such that $|x[n]| \leq A_{max}$ for the intended operating regime. A bipolar (zero-mean) waveform can then be mapped as

$$v[n] = V_{cm} + \rho \frac{\Delta V}{2} \frac{x[n]}{A_{max}}, \quad \tilde{v}[n] = \text{clip}(v[n]; V_{min}, V_{max})$$

where $\text{clip}(u; a, b) = \min(\max(u, a), b)$. This construction makes the *polarity* of $x[n]$ meaningful at the electrical interface (through deviations above/below V_{cm}) while keeping the single-ended voltage nonnegative if $V_{min} \geq 0$. The exact structure covers unipolar waveforms ($x[n] \in [0, A_{max}]$) using a one-sided map,

$$v[n] = V_{min} + \rho \Delta V \frac{x[n]}{A_{max}}, \quad \tilde{v}[n] = \text{clip}(v[n]; V_{min}, V_{max})$$

so that the downstream converter core always sees a signal expressed in physical electrical units and bounded by the hardware swing.

³⁰ M. D'Apuzzo, M. D'Arco, A. Liccardo, and M. Vadursi, “Modeling DAC Output Waveforms”, IEEE Transactions on Instrumentation and Measurement, vol. 59, no. 11, pp. 2854–2862, 2010.

In short-reach optical transmitters, this “headroom + clipping policy” is not a numerical afterthought: it is a performance-relevant design degree of freedom because practical systems deliberately trade off full-scale utilization against clipping distortion and effective resolution, especially in high-speed links where finite DAC swing and bandwidth are limiting factors³¹.

Polarity handling under bounded, nonnegative electrical outputs

The key issue with polarity in IM/DD transmitter chains is not that *electronics* cannot represent signed quantities. Still, that *optical intensity* is nonnegative, whereas many efficient electrical representations (including BPAM-style alphabets and many DSP internal signals) are naturally symmetric around zero. The surrogate resolves this by separating (i) the information-bearing polarity of the electrical waveform (encoded relative to a reference level) from (ii) the absolute electrical swing constraints set by the DAC/driver/modulator interface.

In practice, many high-speed modulator drives are effectively differential, and the modulator response depends on both differential and common-mode components. In segmented or driver-array-based optical transmitters, careful common-mode planning is explicitly required because the modulator transfer can be sensitive to both the sum and difference of the arm voltages; therefore, mapping a “conceptually bipolar” baseband into bounded supply-referenced levels is a fundamental architectural constraint rather than a modeling choice³².

Within this interpretation, the affine remap above has a clear physical meaning:

- **Bipolar** $x[n]$ corresponds to deviations around a chosen operating point (often tied to driver common-mode and modulator bias planning).

Even if the DAC output is constrained to $[0, V_{max}]$ suitable V_{cm} preserves

³¹ M. van den Hout, S. van der Heide, and C. Okonkwo, “Digital Resolution Enhancer Employing Clipping for High-Speed Optical Transmission”, *Journal of Lightwave Technology*, vol. 38, no. 11, pp. 2897–2904, 2020.

³² M. Vanhoecke, A. Aimone, N. Argyris, S. Dris, R. Vaernewyck, K. Verheyen, M. Gruner, G. Fiol, D. Apostolopoulos, H. Avramopoulos, G. Torfs, X. Yin, and J. Bauwelinck, “Segmented Optical Transmitter Comprising a CMOS Driver Array and an InP IQ-MZM for Advanced Modulation Formats”, *Journal of Lightwave Technology*, vol. 35, no. 4, pp. 862–867, 2017.

the sign of $x[n]$ by placing positive samples above V_{cm} and negative samples below (V_{cm} without crossing below zero).

- **Unipolar** $x[n]$ corresponds to direct amplitude coding within $[V_{\text{min}}, V_{\text{max}}]$, appropriate when the upstream symbols are already defined on a nonnegative alphabet (typical for conventional IM/DD PAM intensity levels).

This is why the remap is treated as a *converter interface operation* (inside the DAC block): it sets the physically meaningful reference level and swing that downstream converter-core models (quantization, INL/DNL, reconstruction, noise) should act upon.

Link-level constraints: swing, bias, and “electrical-to-optical feasibility”

Finally, the mapping cannot be designed in isolation from the optical front end. High-swing modulator-driver designs explicitly emphasize that electrical outputs must simultaneously (i) stay within device limits and linearity constraints and (ii) provide adequate modulation index around the chosen bias point of the modulator³³. In a surrogate, enforcing the DAC range and polarity handling at the remap stage is therefore a prerequisite for physically credible downstream modeling: it prevents the simulator from implicitly assuming infeasible drive swings. It ensures that subsequent blocks (driver/modulator) operate within an electrical regime that is consistent with realistic transmitter co-design.

5.3.4 Static transfer function: quantization + code mapping + clipping policy

After electrical-range remapping (Section 5.3.3), the DAC's “static transfer” stage produces the ideal code-voltage staircase that maps a continuous-valued, discrete-time electrical waveform to a finite set of output levels. In the surrogate, this step is viewed as a deterministic mapping (not just “additive quantization

³³ A. H. Ahmed, A. Moznine, D. Lim, Y. Kim, A. Rylyakov, and S. Shekhar, “A Dual-Polarization Silicon-Photonic Coherent Transmitter Supporting 552 Gb/s/Wavelength”, IEEE Journal of Solid-State Circuits, vol. 55, no. 9, pp. 2597–2608, 2020.

noise") because (i) low- to moderate-resolution DACs produce signal-dependent granular distortion, (ii) hard range limits cause overload/clipping distortion, and (iii) the resulting staircase interacts with subsequent reconstruction effects (ZOH + analog filtering) and any upstream pre-emphasis.

Uniform quantizer definition and mid-rise vs. mid-tread conventions

Let $x[n]$ be the (already remapped) discrete-time electrical waveform entering the converter core, constrained to a nominal full-scale range $[V_{min}, V_{max}]$, and let the DAC resolution be N bits with $N_{lev} = 2^N$ output codes. The surrogate implements a *uniform mid-rise* quantizer, in which decision thresholds lie on a uniform grid, and reconstruction levels lie at half-steps between thresholds. The step size (1 LSB) is

$$\Delta = \frac{V_{max} - V_{min}}{N_{lev}}.$$

With a mid-rise convention, the *decision thresholds* are

$$\tau_k = V_{min} + k\Delta, \quad k = 0, 1, \dots, N_{lev},$$

and the *reconstruction levels* are centered in each interval,

$$v_k = V_{min} + \left(k + \frac{1}{2}\right)\Delta, \quad k = 0, 1, \dots, N_{lev} - 1.$$

The corresponding code index (unsigned integer) is obtained by binning/slicing

$$c[n] = \text{clip} \left(\left\lfloor \frac{x[n] - V_{min}}{\Delta} \right\rfloor, 0, N_{lev} - 1 \right)$$

and the quantized output level is

$$x_Q[n] = v_{c[n]} = V_{min} + \left(c[n] + \frac{1}{2}\right)\Delta$$

This formulation directly exposes the two classical distortion mechanisms: *granular distortion* (finite Δ within range) and *overload distortion* (inputs outside $[V_{min}, V_{max}]$ mapped to extreme bins). It also makes explicit how “mid-rise” differs from “mid-tread” quantizers (where one reconstruction level lies exactly at a

symmetry point, such as zero, and thresholds are shifted by $\Delta/2$); both are standard uniform-quantizer conventions, but they imply different endpoint handling and therefore different overload behavior at the rails³⁴.

Code mapping choice inside the DAC block

Since the waveform has already been mapped into the DAC’s physical electrical range in Section 5.3.3, the static transfer function can treat $c[n]$ as a straightforward unsigned code index ($c[n] \in 0, \dots, 2^N - 1$), meaning a direct representation of “which DAC code is selected.” This provides the clearest interface for subsequent non-idealities: INL/DNL are naturally defined as deviations from the ideal staircase indexed by code, and reconstruction effects influence the resulting voltage sequence. Conceptually, alternative binary encodings (such as offset binary versus two’s complement) become relevant only when the digital word format itself needs to be modeled; in this model, that representational issue is intentionally separated from the *physical* converter staircase by an explicit remap step.

Clipping policy and the headroom trade-off

Even with explicit remapping, a DAC model must define *where* clipping is enforced and *how* overload is handled. In the surrogate, the clipping policy is hard saturation to the outermost bins (via the *clip*($\cdot$) operation above), which mirrors the physical reality that the output ladder cannot exceed its supply-referenced range. The practical modeling nuance is not the existence of clipping (it is unavoidable), but the headroom policy that determines how often $x[n]$ approaches the rails: allocating more headroom reduces overload distortion but increases effective Δ utilization loss (and thus raises granular distortion relative to signal swing). This headroom-distortion trade-off is a central design theme in bandwidth-limited optical transmitters where the available electrical swing is constrained, and

³⁴ B. Widrow, I. Kollár, and M.-C. Liu, “Statistical theory of quantization,” IEEE Transactions on Instrumentation and Measurement, vol. 45, no. 2, pp. 353–361, 1996.

waveform design/predistortion choices can shift the balance between clipping and in-band distortion³⁵.

Why does the surrogate use an explicit staircase (not only an AQNM)

For analysis and for closed-form link budgeting, it is common to approximate the quantization error

$$e[n] = x_Q[n] - x[n]$$

as an additive noise term uniformly distributed over $(-\Delta/2, \Delta/2)$ with variance $(\sigma_e^2 = \Delta^2/12)$, under “high-resolution” and “sufficiently busy signal” conditions. This classical statistical framework is valuable because it provides interpretable SNR/ENOB scaling laws and clarifies when quantization behaves like approximately uncorrelated noise.³⁶

However, the additive-noise approximation is not universally reliable, especially for low resolution, strong spectral shaping, or inputs with significant structure, because the error can become signal-dependent and correlated. Rigorous treatments characterize the regimes where the additive-noise model is valid (and where it breaks down), motivating the surrogate’s choice to implement the *exact deterministic* staircase at the waveform level and to treat an additive quantization noise model (AQNM) like” views as optional *derived* interpretations rather than the core model³⁷.

Relevance in high-speed PAM transmitters and consistency with low-resolution practice

The need for clear DAC quantization is particularly important in high-speed IM/DD PAM experiments, where the transmitter bandwidth and converter resolution jointly influence eye opening and BER. Recent IM/DD demonstrations

³⁵ A. Josten, B. Baeuerle, R. Bonjour, W. Heni, and J. Leuthold, “Optical Transmitters without Driver Amplifiers—Optimal Operation Conditions,” *Applied Sciences*, vol. 8, no. 9, art. 1686, 2018.

³⁶ B. Widrow, I. Kollár, and M.-C. Liu, “Statistical theory of quantization,” *IEEE Transactions on Instrumentation and Measurement*, vol. 45, no. 2, pp. 353–361, 1996.

³⁷ D. Marco and D. L. Neuhoff, “The validity of the additive noise model for uniform scalar quantizers,” *IEEE Transactions on Information Theory*, vol. 51, no. 5, pp. 1739–1755, 2005.

explicitly use very low-resolution DACs (e.g., 4-bit and 3-bit) and rely on pre-equalization and noise-shaping techniques to mitigate in-band impairments, showing that quantization is not merely a secondary noise source but a primary transmitter constraint that must be addressed at the appropriate point in the signal chain³⁸.

In even more aggressive low-resolution settings, the spectrum of quantization distortion and its redistribution by transmitter-side processing becomes part of the design space, emphasizing the distinction between (i) the ideal staircase (this section) and (ii) the conditioning/reconstruction blocks that influence how the staircase appears in time and frequency³⁹.

DAC vs. ADC quantization in the surrogate

Although the mathematics of uniform quantization is shared between DACs and ADCs, their system roles differ. DAC quantization shapes the transmitted waveform before the driver or modulator, affecting transmitter conditioning, reconstruction, and subsequent analog or optical nonlinearities. ADC quantization operates on a waveform that has already accumulated channel noise or distortion and typically follows anti-aliasing filtering and sampling constraints. For this reason, the thesis treats DAC quantization primarily as a transmitter impairment that affects the launched waveform. In contrast, ADC quantization is viewed as a receiver limitation that affects observability and DSP performance, even when the underlying scalar quantizer equations appear similar⁴⁰.

5.3.5 Static nonlinearities beyond quantization: INL/DNL (definitions, injection point, parameters)

³⁸ M. Yin, D. Zou, W. Wang, F. Li, and Z. Li, “Transmission of a 56 Gbit/s PAM-4 signal with low-resolution DAC and pre-equalization only over 80 km fiber in C-band IM/DD systems for optical interconnects,” *Optics Letters*, vol. 46, no. 22, pp. 5615–5618, 2021.

³⁹ J. Jiang, M. Yin, D. Yang, and F. Li, “Generation of high-speed PAM-4 signal with 3-bit DAC enabled by CRD-NS in optical interconnect,” *Optics Express*, vol. 31, no. 13, pp. 21153–21160, 2023.

⁴⁰ B. Widrow, I. Kollár, and M.-C. Liu, “Statistical theory of quantization,” *IEEE Transactions on Instrumentation and Measurement*, vol. 45, no. 2, pp. 353–361, 1996.

Section 5.3.4 defined the *ideal* static transfer of the converter core, uniform quantization, code mapping, and a specified clipping or headroom policy, meaning the mapping from a limited electrical input range to an *ideal* set of evenly spaced reconstruction levels. In reality, however, the DAC static transfer characteristic deviates from this ideal because the physical code steps are not perfectly uniform (for example, element mismatch, limited settling of code-dependent static errors, and static gain or offset variation across codes). These deviations are usually described by *differential nonlinearity (DNL)* and *integral nonlinearity (INL)*, which together represent *static* code-dependent distortions that are different from quantization noise and from *dynamic* effects such as ZOH/reconstruction, glitching, and analog bandwidth limitations.

Let the DAC have (N) bits and $(M = 2^N)$ codes. Denote the ideal least significant bit (LSB) step by

$$\Delta = \frac{V_{max} - V_{min}}{M}$$

And the *ideal* transition (decision) levels by

$$T_k^{\text{ideal}} = V_{min} + k\Delta, \quad k = 0, \dots, M$$

In an *actual* static DAC, the code widths $(T_{k+1} - T_k)$ are not all equal. The DNL of code k is typically defined (in LSB units) as

$$\text{DNL}[k] = \frac{(T_{k+1} - T_k)}{\Delta} - 1, \quad k = 0, \dots, M - 1,$$

so $\text{DNL}[k] = 0$ corresponds to an ideal code width. The INL expresses the cumulative deviation of the transfer characteristic from the ideal straight line, commonly referenced to a chosen endpoint or best-fit line; in an endpoint-referenced form, it can be expressed (again in LSB units) via the deviation of transition levels from the ideal:

$$\text{INL}[k] = \frac{T_k - T_k^{\text{ideal}}}{\Delta}.$$

A key structural point is that INL and DNL are not independent: INL is a cumulative manifestation of local step-width errors (DNL), so modeling that starts from code-width perturbations naturally induces an INL curve once a reference convention and endpoint constraints are enforced. This viewpoint is also aligned with practical static characterization approaches, where the goal is to recover (or

emulate) the INL curve over all codes under realistic measurement-time constraints⁴¹.

Injection point in the surrogate. In the present DAC surrogate, static nonlinearities are applied within the *converter-core* static transfer stage, i.e., *after* (i) the waveform has been brought into the DAC’s electrical operating range (remap) and (ii) the *ideal* quantizer/code mapping has been defined, but *before* time-domain reconstruction mechanisms (ZOH), glitch generation, clock-jitter sampling perturbations, and analog output filtering. This ordering is intentional: INL/DNL are *static transfer* effects (code-dependent amplitude errors and unequal code widths) and must therefore shape the discrete-time code-to-voltage relationship *prior* to reconstruction and bandwidth-limited smoothing. Conceptually, the surrogate implements a perturbed transition-level set $\{T_k\}$ (or an equivalent LUT representation) so that the same input sample can map to a code (and hence an output level) that reflects the DAC’s static imperfections; subsequent stages then propagate those static errors through the continuous-time reconstruction and output path.

Parameterization and modeling strategy. The implementation is structured to support both (a) measurement-informed characterization and (b) statistical generation of plausible static nonlinearity profiles for “what-if” design studies and ML training. Practically, this is realized by representing the static transfer through a *monotone* edge/transition-level LUT $\{T_k k\}_{k=0}^M$. In a statistical mode, the model can perturb code widths (DNL-driven construction) with controlled dispersion (*e.g.*, σ_{DNL} in LSB units), enforce monotonicity (no negative widths), and renormalize the final edge set so that the overall span remains consistent with the converter’s full scale. From such perturbed widths, an INL curve emerges as the cumulative deviation relative to the ideal line, with the chosen referencing convention enforced by construction. In a measurement-informed mode, the LUT (or an INL curve from

⁴¹ P. Daponte, L. De Vito, G. Iadarola, and S. Rapuano, “A reduced-code method for Integral Nonlinearity testing in DACs”, *Measurement*, vol. 182, art. 109764, 2021.

characterization) can be injected directly so the surrogate reproduces device-specific static behavior. This separation is important for avoiding double counting: large-scale gain or offset adjustments are handled by the remap stage (Section 5.3.3), while INL/DNL here represent residual code-dependent deviations around the calibrated operating point.

Why this matters at the system level (and for ML). In high-speed PAM/BPAM transmitters, INL/DNL appear as deterministic, level-dependent amplitude distortion that can (i) distort intended symbol spacing, (ii) generate harmonic and intermodulation components when driven by multilevel waveforms, and (iii) produce pattern-dependent effective impairments when combined with bandwidth limits and downstream nonlinearities. From a surrogate perspective, these effects are not “optional detail”: they alter the impairment structure the optimizer must learn to compensate. Analytical end-to-end analyses that explicitly account for DAC mismatch/non-ideality show that transmitter distortions lead to observable penalties such as EVM and constellation/level warping, underscoring the importance of placing static DAC imperfections in the correct part of the chain when using the surrogate for credible trade-off exploration or for training data-driven compensators⁴². For ML workflows specifically, the LUT-based representation offers a clean abstraction: the static transfer can be exposed as a compact parameter set (or stored as metadata) for supervised learning, inverse modeling, and robustness studies (e.g., training across a range of INL/DNL profiles rather than a single “ideal DAC”).

5.3.6 Dynamic reconstruction behavior: ZOH + output network (and where glitch fits)

After the static transfer stage (code mapping, quantization, clipping, and optional INL/DNL), a high-speed DAC still exhibits *dynamic* behavior that is not

⁴² A. Nazarathy and I. Tomkos, “All-Analytic Statistical Modeling of Constellations in (Optical) Transmission Systems Driven by High-Speed Electronic Digital to Analog Converters Part I: DAC Mismatch Statistics, Metrics, Symmetries, Error Vector Magnitude”, *Photonics*, vol. 11, no. 8, art. 747, 2024.

captured by a memoryless mapping. In practice, the analog output is shaped by (i) the sample-and-hold reconstruction (modeled here as a zero-order hold, ZOH), (ii) the output network/reconstruction filtering (finite bandwidth and settling), and (iii) switching transients (“glitches”) that occur at code transitions. These effects introduce *memory* and *spectral shaping* that directly impact the transmitted waveform, the eye opening, and the spectrum observed at the driver input.

In our DAC pipeline, these effects are applied in the physically consistent order implemented in the code: first static transfer, then ZOH, then glitch injection, then analog output filtering (and subsequent analog impairments/noise, if enabled). This ordering matters because glitches occur on the held waveform during switching events, and the output network partially filters both the held steps and the transient components.



Figure 5.x the injection point of dynamic non-idealities

ZOH (piecewise-hold reconstruction)

A ZOH represents the DAC’s sample-and-hold behavior: each discrete-time sample is held constant for one DAC clock period $T_s = 1/F_s$. In continuous time, the reconstructed waveform can be written as

$$x_{\text{ZOH}}(t) = \sum_{n=-\infty}^{\infty} x[n] \Pi\left(\frac{t - nT_s}{T_s}\right),$$

where $\Pi(\cdot)$ is a rectangular pulse of duration T_s . Equivalently, the ZOH is a convolution with a rectangular pulse, which leads to the well-known frequency response

$$H_{\text{ZOH}}(f) = T_s \text{sinc}(\pi f T_s) e^{-j\pi f T_s},$$

creating the characteristic sinc droop across the baseband and periodic spectral

images around multiples of F_s . This is a purely deterministic “reconstruction” effect: even with an ideal static transfer, ZOH alone reshapes the spectrum and sets the baseline against which reconstruction filtering and pre-emphasis operate⁴³.

Implementation in our code: The function realizes the hold operation in discrete time by expanding the waveform onto a finer time grid: each sample is repeated by an integer factor L_{ZOH} (implemented via a repeat/hold operation), and the effective sampling rate is updated accordingly. This is a modeling convenience: it produces a staircase waveform with sufficient time resolution for (i) short glitch pulses and (ii) subsequent analog filtering, without requiring a fully continuous-time simulator.

Output network/reconstruction filtering (finite bandwidth + settling)

Real DAC outputs are shaped by the combined response of the *output buffer, package/interconnect parasitics, bondwires, PCB traces, and the load*. A widely used system-level abstraction is to treat this path as a (dominantly) linear time invariant (*LTI*) *reconstruction network* with impulse response $h_{\text{out}}(t)$ and frequency response $H_{\text{out}}(f)$. The resulting analog waveform becomes

$$x_{\text{out}}(t) = (x_{\text{ZOH}} * h_{\text{out}})(t), \quad X_{\text{out}}(f) = X(f)H_{\text{ZOH}}(f)H_{\text{out}}(f).$$

From a communications viewpoint, this stage is responsible for bandwidth limitation and time-domain smoothing, but also for deterministic ISI (channel memory) when the output response is not negligible compared to the symbol interval. This means the output buffer, interconnects, and load are assumed not to introduce nonlinear distortion (no compression, clipping, or AM–AM/AM–PM effects), and all distortions are modeled as linear filtering effects.

Therefore, reconstruction filtering is not merely “cleanup”; it changes the effective transmit pulse shape and must be co-designed with DSP (e.g., pre-emphasis/equalization) and with the clipping/headroom policy.

⁴³ A. V. Oppenheim and R. W. Schaffer, “*Discrete-Time Signal Processing*”, 3rd ed. Pearson, 2010. ISBN: 978-0131988422.

Implementation in our code: This block implements an analog-style low-pass reconstruction stage using a parameterized digital filter applied to the ZOH-expanded waveform. The current implementation supports configurable cutoff and order, and is intended as the *first-order surrogate* of $H_{\text{out}}(f)$ when detailed package models are unavailable. Finite settling and waveform distortion arising from limited output dynamics are precisely the class of effects this block targets at the system level⁴⁴.

A complementary perspective is that finite-settling and time-domain DAC specifications (rise/fall behavior, overshoot/settling, and related spectral consequences) can be interpreted as constraints on $h_{\text{out}}(t)$ and are therefore naturally modeled as part of the reconstruction network rather than the static transfer map⁴⁵.

Where glitch fits: transient switching component tied to code transitions

Beyond ZOH and LTI output filtering, high-speed DACs exhibit *switching transients*: short-duration pulses generated when the DAC changes code. These arise from current-steering mismatch, switching skew, charge injection, dynamic element selection, and imperfect return-to-zero behavior in internal switching networks. In the time domain, a compact and simulation-friendly representation is to model glitch as an additive transient superimposed on the reconstructed waveform:

$$x(t) = x_{\text{ZOH}}(t) + \sum_k a_k p(t - t_k),$$

where t_k are the transition instants (code changes), $p(t)$ is a short pulse shape (e.g., half-sine or Gaussian), and a_k is an amplitude/energy coefficient that can depend

⁴⁴ M. D’Apuzzo, M. D’Arco, A. Liccardo, and M. Vadursi, “Analyzing DAC waveform distortion due to finite settling time”, in *Proc. IEEE Instrumentation & Measurement Technology Conference (I2MTC)*, pp. 1398–1403, 2010.

⁴⁵ E. Balestrieri, “DAC Time-Domain Specifications Toward St...surement”, *IEEE Transactions on Instrumentation and Measurement*, vol. 57, no. 7, pp. 1290–1297, Jul. 2008.

on the code step size (Δx_k), the direction of the transition, and (optionally) stochastic effects such as dynamic element matching (DEM).

This modeling choice aligns with how glitch is treated in spur/SFDR analysis: glitch energy produces *spectral spurs and harmonic distortion*, and its impact depends strongly on both the transition statistics and the subsequent filtering⁴⁶. For binary-weighted current-steering DACs, explicit circuit-level and mixed-signal studies show that reducing glitch energy can translate directly into improved spurious-free dynamic range (SFDR), motivating a system-level knob that controls glitch amplitude/energy and its occurrence⁴⁷. Moreover, DEM strategies can change how glitch energy manifests (e.g., by redistributing transition events in time and across elements), which justifies supporting stochastic or transition-dependent glitch activation in the surrogate⁴⁸.

Implementation in our code: The current block is intentionally written as a *general injection framework*:

- It detects code transitions on the ZOH waveform (so the transient aligns with the held steps).
- It generates a short pulse kernel $p(t)$ (currently supported pulse families include *half-sine* and *Gaussian* forms, consistent with common short-pulse approximations used in mixed-signal modeling).
- It scales the pulse per transition (deterministically and/or stochastically) and adds it to the waveform.

⁴⁶ H.-Y. Huang, X.-Y. Chen, and T.-H. Kuo, “A 10-GS/s NRZ/Mixing DAC With Switching-Glitch Compensation Achieving SFDR > 64/50 dBc Over the First/Second Nyquist Zone,” IEEE Journal of Solid-State Circuits, vol. 56, no. 10, pp. 3145–3156, 2021.

⁴⁷ F. Chou and C. Hung, “Glitch Energy Reduction and SFDR Enhancement Techniques for Low-Power Binary-Weighted Current-Steering DAC,” IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 24, pp. 2407–2411, 2016.

⁴⁸ W. Liu, T. Zhang, H. Zhang, H. Tang, and L. Zhang, “A glitch energy reduction method with low complexity dynamic element matching for data converters,” IEICE Electronics Express, vol. 20, 20220529, 2023.

Analog output network and reconstruction filtering

After the ideal sample-and-hold/ZOH stage, the DAC output is not observed directly at the converter pins. Instead, it propagates through the converter output buffer, package/PCB/load parasitics, and (optionally) an explicit reconstruction network. In the surrogate, this post-ZOH analog path is modeled as a continuous-time LTI system with transfer function $H_{\text{out}}(s)$, so that the reconstructed voltage is

$$v_{\text{out}}(t) = (v_{\text{ZOH}} * h_{\text{out}})(t),$$

where $v_{\text{ZOH}}(t) = \sum_n x[n] \Pi\left(\frac{t-nT_s}{T_s}\right)$ is the piecewise-constant hold waveform and $h_{\text{out}}(t)$ is the impulse response of the analog output network. In the frequency domain, the overall reconstruction seen by the downstream driver is conveniently expressed as

$$V_{\text{out}}(f) = X_d(e^{j2\pi f T_s}) H_{\text{ZOH}}(f) H_{\text{out}}(j2\pi f),$$

$$\text{with the ZOH envelope } H_{\text{ZOH}}(f) = T_s \text{sinc}(fT_s) e^{-j\pi f T_s},$$

which already introduces a deterministic magnitude droop and spectral imaging. Therefore, $H_{\text{out}}(s)$ is not a cosmetic “smoothing” element: it determines (i) how strongly the image bands are suppressed, (ii) the in-band magnitude/phase distortion that translates into memory (ISI-like) effects in PAM waveforms, and (iii) the effective analog noise bandwidth at the DAC output. This treatment is consistent with measurement-oriented DAC modeling where settling/bandwidth limitations are represented by low-order analog networks driven by code transitions, yielding predictable spectral consequences at high rates⁴⁹. In high-speed optical-transmitter contexts, the same principle is applied by extracting/using a composite frequency response of the DAC plus its analog output path in system

⁴⁹ M. D’Apuzzo, M. D’Arco, A. Liccardo, and M. Vadursi, “Analyzing DAC waveform distortion due to finite settling time”, IEEE Instrumentation & Measurement Technology Conference (I2MTC), pp. 352–356, 2010.

simulations, because this response bounds the usable modulation bandwidth and directly impacts eye-diagram quality⁵⁰.

Different filter choices correspond to distinct physical hypotheses about the output network and yield distinct distortion signatures. A minimal “buffer+load” approximation is a first-order low-pass,

$$H_{\text{out}}(s) = \frac{1}{1 + s/\omega_c},$$

which captures a dominant pole (slew/settling time constant) and yields smooth but strongly frequency-dependent group delay near ω_c . A more realistic representation for ringing/incomplete settling is a second-order section

$$H_{\text{out}}(s) = \frac{\omega_0^2}{s^2 + 2\zeta\omega_0 s + \omega_0^2},$$

where $\zeta < 1$ models under-damped peaking and time-domain overshoot, while $\zeta > 1$ models over-damped sluggish settling; this is precisely the regime where “settling error at T_s ” becomes a rate-dependent impairment because residual transients leak into the next symbol interval⁵¹. When an explicit reconstruction filter is intended (beyond parasitics), classical low-pass families provide design degrees of freedom: *Butterworth filters* offer maximally flat passband magnitude ($|H(j\omega)|^2 = 1/(1 + (\omega/\omega_c)^{2N})$) at the cost of non-linear phase; Bessel filters preserve near-constant group delay over the occupied band (often preferable when eye opening is limited by dispersion of the electrical phase response rather than by out-of-band images); Chebyshev/elliptic prototypes increase transition steepness but introduce ripple that can translate into amplitude modulation and increased deterministic ISI. In practice, the “best” choice is dictated by whether the objective is primarily image suppression or waveform fidelity (EVM/eye opening), and by

⁵⁰ C. Schmidt, C. Kottke, R. Freund, F. Gerfers, and V. Jungnickel, “*Digital-to-analog converters for high-speed optical communications using frequency interleaving: impairments and characteristics*”, Optics Express, vol. 26, no. 6, pp. 6758–6770, 2018.

⁵¹ M. D’Apuzzo, M. D’Arco, A. Liccardo, and M. Vadursi, “*Analyzing DAC waveform distortion due to finite settling time*”, IEEE Instrumentation & Measurement Technology Conference (I2MTC), pp. 352–356, 2010.

how aggressively the downstream chain relies on digital pre-emphasis to invert the composite response. A representative example of reconstruction-filter co-design shows that the LPF order and cutoff must be selected jointly with sampling/oversampling choices to balance in-band distortion against image suppression, exactly the trade-off the surrogate must expose when exploring bandwidth-limited PAM transmitters⁵².

5.3.7 Noise model set and parameterization strategy (non-quantization noises)

In addition to deterministic impairments (static transfer, INL/DNL, ZOH, and reconstruction), a DAC-driven transmitter exhibits *stochastic* distortions that cannot be captured by a purely static mapping. In the surrogate, we separate these effects into two physically distinct classes: (i) *amplitude-domain noise* (additive voltage/current perturbations generated by the converter core and its analog output path), and (ii) *timebase noise* (clock/aperture jitter that perturbs the sampling/update instants). This separation is consistent with behavioral DAC modeling practice, where the converter is represented as an ideal sampled-data operator plus a compact set of additive and timing noise mechanisms that can be parameterized from measurable specifications (e.g., noise density, SNR/SINAD, jitter RMS, SFDR)⁵³.

Noise taxonomy beyond quantization (behavioral viewpoint)

While quantization noise was addressed in the static transfer section, high-speed DACs add extra noise sources that are typically grouped into: (a) *output buffer/load thermal noise*, (b) *reference and bias network noise*, (c) *power supply-induced noise* (including ripple-to-output coupling), (d) *flicker (1/f) and low-frequency noise* that can upconvert through switching and interleaving, and (e) *digital switching/substrate coupling noise*, which appears as either an

⁵² S. Cheon, S. Kweon, A. Abera, M. Je, and S. Ha, “Optimal Parameter Design of DAC-Based Sinusoidal Signal Generators for Electrical Impedance Spectroscopy”, *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–11, 2022.

⁵³ H. Yu, T. Michalka, M. Larbi, and M. Swaminathan, “Behavioral Modeling of Tunable I/O Drivers With Preemphasis Including Power Supply Noise,” *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 28, pp. 233–242, 2020.

approximately wideband floor or as partially periodic spurs, depending on architecture and activity. Behavioral models for communication-grade current-steering DACs generally represent these effects at the system level as *additive noise sources* (white and/or colored) injected at the electrical output, optionally supplemented by spur terms when interleaving or switching periodicities are significant⁵⁴.

A pragmatic implication for a *digital-surrogate* workflow is that the above mechanisms need not be modeled at device-physics granularity from day one. Instead, they can be expressed through a small set of identifiable parameters(e.g., an equivalent one-sided voltage-noise PSD $S_v(f)$, one or more corner frequencies for colored components, and spur magnitudes/frequencies), then refined later by fitting to measured output spectra or SINAD/SFDR trends. This “spec-to-model” approach is also used in optical-link analyses, where DAC/ADC noise sources are modeled via effective SNR/SINAD terms rather than transistor-level descriptions, enabling system-level prediction and optimization⁵⁵. A broader statistical modeling perspective in optical communications similarly emphasizes *parameterized hardware-impairment models* (noise/jitter/phase and amplitude imperfections) as the correct abstraction level when the goal is accurate performance prediction and optimization under realistic hardware constraints⁵⁶.

Why we currently bundle several mechanisms into a thermal-noise term

At the current surrogate maturity stage, we implement an effective broadband additive noise term (referred to as “thermal noise” in the code) as the *first-order* representation of the DAC’s non-quantization amplitude noise. The intent is not to claim that all mechanisms are literally Johnson noise, but to provide a controlled

⁵⁴ W. Dghais, T. Cunha, and J. Pedro, “Behavioral model for high-speed digital buffer/driver”. *Proc. INMMiC*, 110–113, 2010

⁵⁵ T. Ye, X. Su, C. Yang, J. Li, Z. Tao, H. Nakashima, “Noise analysis for communication systems using high-speed DAC and ADC”. *ECOC* 2022.

⁵⁶ M. Nazarathy and I. Tomkos, “All-Analytic Statistical Modeling of Constellations in (Optical) Transmission Systems Driven by High-Speed Electronic Digital to Analog Converters Part I: DAC Mismatch Statistics, Metrics, Symmetries, Error Vector Magnitude”. *Photonics*, 11(8), 747, 2024.

equivalent in-band noise floor that captures the aggregate effect of multiple wideband contributors (buffer + load + mixed-signal coupling) in a way that is easy to calibrate and stable for ML dataset generation. This “equivalent-noise” approach aligns with behavioral DAC modeling, in which several internal noise mechanisms are collapsed into an output-referred additive term unless a specific spur/corner behavior is under study⁵⁷.

Mathematically, the surrogate uses an output-referred noise model

$$x_{\text{noisy}}[n] = x[n] + w[n], \quad w[n] \sim \mathcal{N}(0, \sigma_v^2),$$

where σ_v^2 is parameterized from physical or measured quantities. In the most physically grounded form (used in the current implementation), σ_v is computed from the Johnson–Nyquist relation over an effective noise bandwidth (B),

$$\sigma_v^2 = 4k_B T R B,$$

with k_B the Boltzmann constant, T the absolute temperature, and R an equivalent load resistance (e.g., 50 Ω environment). If the post-DAC analog path is modeled as an LTI filter $H_{\text{out}}(f)$, the more general form is

$$\sigma_v^2 = \int_{-\infty}^{\infty} S_v(f) |H_{\text{out}}(f)|^2 df,$$

which makes explicit that “thermal noise” in system simulations is ultimately an *output-referred PSD integrated through the reconstruction network*. This is the reason bundling is convenient at this stage: the single parameter $S_v(f)$ (or equivalently σ_v in-band) can later be split into multiple contributors once measurements justify the added complexity.

⁵⁷ W. Dghais, T. Cunha, and J. Pedro, “Behavioral model for high-speed digital buffer/driver”. *Proc. INMMiC*, 110–113, 2010.

Clock/aperture jitter as a timebase-noise mechanism (and why it matters)

Unlike additive amplitude noise, DAC clock jitter perturbs the time at which the held output is updated. Let the ideal update instants be ($t_n = nT_s$). With clock jitter,

$$t_n = nT_s + \delta t_n,$$

and the sampled-data value is effectively drawn from the continuous-time waveform at the perturbed instant. For sufficiently small jitter (relative to the waveform curvature), a first-order expansion gives the canonical linearized model:

$$x(t_n) \approx x(nT_s) + \left. \frac{dx(t)}{dt} \right|_{t=nT_s} \delta t_n \Rightarrow e_j[n] \approx x'(nT_s) \delta t_n$$

so jitter-induced error is *signal-dependent* (proportional to slope) and therefore grows with occupied bandwidth and high-frequency content. A classic closed-form result for a sinusoid at frequency (f) yields the widely used jitter-limited SNR trend

$$\text{SNR}_j \propto \frac{1}{(2\pi f \sigma_t)^2}$$

where σ_t is RMS jitter⁵⁸.

For broadband modulations (PAM/QAM/OFDM), the same principle holds: jitter maps timing uncertainty into an effective noise term whose variance depends on the waveform's RMS slope, and it can be expressed as an EVM/SNR penalty under appropriate assumptions. A particularly useful system-level result is a *closed-form BER/EVM degradation model* in high-speed wireline links, which is directly transferable to the DAC-front-end viewpoint here because the underlying mechanism (timebase perturbation of a bandlimited waveform) is the same⁵⁹. When EVM is used as a bridge metric, established relationships between EVM and SNR

⁵⁸ N. Kurosawa, H. Kobayashi, K. Maruyama, H. Sugawara, and K. Kobayashi, "Sampling clock jitter effects in digital-to-analog converters". *Measurement*, 31(3), 187-199, 2002.

⁵⁹ M. Oveisi and P. Heydari, "A Study of BER and EVM Degradation in Digital Modulation Schemes Due to PLL Jitter and Communication-Link Noise". *IEEE Transactions on Circuits and Systems I: Regular Papers*, 69(8), 3402 – 3415, 2022.

in practical receivers provide a consistent way to convert a jitter-induced EVM increase into an SNR penalty for link budgeting and optimizer objectives⁶⁰.

In the current surrogate, clock jitter is implemented explicitly as a *time perturbation sequence* δt_n (with random and optional deterministic components), then applied using one of three strategies: (i) the *linearized slope-noise* model $x[n] + x'[n]\delta t_n$ (accurate for small jitter), (ii) *edge-shift* behavior (capturing sample-update displacement effects), or (iii) full *time-warp resampling* (more expensive but closer to a direct sampled-time interpretation). This makes jitter a first-class, independently parameterized impairment, rather than an arbitrary additive noise term, because it interacts differently with spectrum shaping, reconstruction filtering, and any downstream equalization/DPD processes in unique ways.

Why are these noise models essential for ML/NN-DPD workflows?

A key motivation for modeling non-quantization noise and jitter explicitly is that learning-based predistortion and optimization are highly sensitive to training-domain realism. If the surrogate generates “clean” waveforms that omit the stochastic impairments present in the lab, an NN-based DPD can converge to an inverse that is optimal only for the synthetic domain, leading to a deployment mismatch (domain gap) and degraded real-system performance. Work on robust NN-based DPD explicitly addresses *hardware-impairment robustness* and shows that incorporating impairment variability during training (or using domain-randomization strategies) improves generalization⁶¹. In coherent optical contexts, robust NN-DPD frameworks likewise emphasize that the predistorter must remain

⁶⁰ H. A. Mahmoud and H. Arslan, “Error Vector Magnitude to SNR Conversion for Nondata-Aided Receivers”. *IEEE Transactions on Wireless Communications*. 8(5), 2009.

⁶¹ P.J. Freire, A. Napoli, B. Spinnler, N. Costa, S.K. Turitsyn, J.E. Prilepsky, “Neural Networks-Based Equalizers for Coherent Optical Transmission: Caveats and Pitfalls”, *IEEE Journal of Selected Topics in Quantum Electronics*, 28(4), 7600223, 2022.

effective under realistic transmitter/receiver distortions and noise conditions rather than idealized ones⁶².

More broadly, learning-based DPD methodologies (including “augmented learning” strategies) treat the transmitter as a nonlinear, possibly memory-bearing system whose effective inverse should be learned *in the presence of realistic distortions and noise*, because the optimization objective is ultimately performance under operating conditions rather than noiseless functional inversion⁶³. Finally, hardware-robustness discussions of hardware robustness in NN-based optical DSP highlight the same message from a system standpoint: synthetic training must include representative impairments (converter limitations, noise, timing uncertainty) to avoid brittle models that fail when moved from simulation to hardware⁶⁴.

5.3.8 Metadata and calibration hooks for ML/optimization

A central design choice of the proposed surrogate is that the DAC block is not only a numerical waveform transformer, but also a measurement and traceability point: it produces (i) a physically consistent electrical waveform after conversion modeling, and (ii) a structured record that makes each run *reproducible, searchable, and learnable*. This section formalizes the metadata philosophy adopted in the implementation: every transformation in the DAC module is accompanied by *machine-readable descriptors* that capture operating conditions, impairment configuration, and provenance. This is essential for two later uses that go beyond conventional link simulation: *calibration* against measurements (so the surrogate remains anchored to hardware behavior) and *ML/optimization* workflows (so

⁶² M. Sena, M.S. Erkilinc, T. Dippon, B. Shariati, R. Emmerich, J.K. Fischer, “*Bayesian Optimization for Nonlinear System Identification and Pre-Distortion in Cognitive Transmitters*”. Journal of Lightwave Technology, 39(15), 5008–5020, 2021.

⁶³ D. Wang, M. Aziz, M. Heloui, & F. Ghannouchi, “*Augmented Real-Valued Time-Delay Neural Network for Compensation of Distortions and Impairments in Wireless Transmitters*”. IEEE Transactions on Neural Networks and Learning Systems, 30(1), 242-254, 2019.

⁶⁴ T. Sasai, M. Nakamura, E. Yamazaki, and Y. Kisaka, “*Toward Hardware Robustness for Neural Network-Based Digital Signal Processing in Optical Communications*,” Optics Express, vol. 28, no. 11, pp. 15780–15799, 2020.

training data can be generated with correct condition labels and can be audited/filtered when performance is analyzed).

A. Two-layer record: waveform snapshots vs. metadata descriptors

The codebase follows a two-layer logging strategy.

1. **Waveform snapshot layer (signal traceability).** At the end of each processing stage, a snapshot of the evolving waveform is retained so that the end-to-end receiver can be trained or validated not only on the final received samples, but also on intermediate representations (e.g., post-remap, post-quantization, post-INL/DNL, post-ZOH, post-output filtering, post-noise injection). This enables “probe-point” supervision: ML models can be trained on *paired inputs/outputs* at specific boundaries (e.g., “what the DAC core produced” vs. “what the output network delivered”), which is critical when the target is a learned compensator (DPD) or an optimizer that must attribute performance degradation to specific impairment classes.

2. **Metadata layer (condition tags, impairment labels, and provenance).** In parallel, the DAC block emits a structured descriptor that records *what was applied and under which conditions*, including both configuration (what the run intended to do) and realized statistics (what the run actually produced). This layer is the backbone for dataset governance: it allows one to query, for example, “all realizations with clipping ratio <1% and RMS jitter in [0.3, 0.6] ps,” or “all runs where INL was enabled but DNL disabled,” and then link those subsets to BER/EVM outcomes.

B. What is stored, at what granularity, and why

The metadata is most valuable when it is block-structured and granularity-aware, i.e., it separates (i) run-level provenance, (ii) block-level configuration, and (iii) stage-level outcomes. A scientifically useful schema for the DAC module, therefore, includes:

- **Run/scenario identifiers (reproducibility keys).** A scenario ID (or deterministic run tag), seed policy and seed value(s), and a compact representation of the configuration state (e.g., a parameter hash or profile

name). This ensures results can be regenerated *without relying on execution history*. It also enables external logging (tables of experiments, ablation studies, training datasets) to point to a unique run identity.

- **Domain and unit tags (interoperability).** Since the DAC is the digital-to-electrical boundary, each stored item should be labeled with its **domain** and **units** (e.g., dimensionless normalized sample stream vs. volts at the DAC output). This prevents downstream ambiguity, especially when the same waveform is reused for different experiments (e.g., comparing electrical conditioning vs. optical modulation). In a multi-domain surrogate, unit tags serve as correctness constraints rather than just cosmetic labels.

- **Stage enable/disable flags (ablation and interpretability).** Each impairment mechanism is logged by explicit on/off status (e.g., quantization, INL, DNL, ZOH, output filtering, thermal noise, clock jitter, glitch model). This supports controlled studies and ML dataset curation: supervised training benefits from condition labels (“impairment present/absent”), and scientific reporting benefits from unambiguous ablation tables.

- **Parameter values and model forms (scientific interpretability).** When an impairment is enabled, the metadata should record the model form and the minimal defining parameters (e.g., number of bits and full-scale range for quantization; INL/DNL profile choice and scale; jitter standard deviation; thermal noise bandwidth/temperature and derived variance; filter order/cutoff for reconstruction network). The goal is not to duplicate the full parameter tree, but to store the subset needed to interpret results and to reproduce the waveform transformation.

- **Outcome statistics (link-level relevance).** Alongside configuration, the DAC block should log post-stage statistics that are meaningful for later performance analysis and optimization: RMS/peak voltage, mean (DC) level, clipping incidence, quantization step size, estimated ENOB proxies (if computed), and any bandwidth-related measures (e.g., post-filter power distribution or spectral roll-off indicators

when available). These values provide “features” for learning pipelines and sanity checks for calibration.

C. Calibration hooks: separating “knobs” from “identified parameters.”

A key architectural goal is to keep a strict separation between:

- **Nominal/datasheet parameters** (static design-time constraints),
- **Calibration parameters** (identified from measurements), and
- **Operational knobs** (deliberately tuned during experiments/optimization).

Within the DAC block, calibration hooks matter most in two places:

1. **Timebase anchoring at the converter boundary.** The DAC time resolution (sampling rate after match-Fs) defines the physical time grid on which reconstruction, jitter, glitch, and output filtering act. Treating this as a DAC-owned responsibility makes calibration meaningful: measured transfer characteristics of the electrical output network (or driver input response) can be referenced to a single timebase without “hidden resampling” elsewhere.

2. **Output-path identification and reconstruction network tuning.** The reconstruction filter/output network is exactly the kind of element that can be *fitted* from measurement (step response, frequency response, or S-parameter-derived magnitude/phase). The code structure benefits from representing this as a calibration-fed LTI block whose parameters can be updated without rewriting the rest of the DAC chain. The metadata must therefore retain the calibration version/tag (e.g., the fit used) to ensure traceability when results change after recalibration.

This calibration-aware design also enables credible optimization: Bayesian optimization, gradient-free search, and ML/RL-based tuning require that “what is being optimized” be clearly separated from “what is fixed by hardware

identification.” Without that separation, optimizers may learn compensations that are not transferable to the lab⁶⁵.

D. ML/optimization relevance: dataset traceability and impairment-aware learning

The surrogate’s ML value does not come from storing “more data,” but from storing the right labels at the right boundaries. In DAC-centric learning problems, especially NN-based, the dominant risk is mismatch: a model trained on synthetic data that does not reflect the actual impairment mixture (or the correct injection point) may learn the wrong inverse mapping and fail to generalize. The DAC metadata strategy directly addresses this by:

- Enabling impairment-conditional training (train distinct models for distinct impairment regimes, or train a single conditional model with impairment tags as inputs).
- Supporting curriculum-style datasets (start with linear/noiseless DAC, then progressively enable quantization, INL/DNL, reconstruction filtering, jitter, and noise).
- Providing auditability: when a trained model fails, the training set can be filtered by scenario IDs and impairment labels to isolate which conditions are underrepresented or inconsistent.
- Allowing block-local optimization: because DAC parameters and conditioning filters are logged and controlled within one block boundary, optimizers can adjust them while maintaining a consistent interpretation of results across runs.

5.4 Driver and electrical front-end (post-DAC analog path)

In the proposed surrogate chain, the driver is the first explicitly analog electrical subsystem after the DAC, making it the point where the modeled DAC waveform

⁶⁵ M. Sena, M.S. Erkilinc, T. Dippon, B. Shariati, R. Emmerich, J.K. Fischer, “*Bayesian Optimization for Nonlinear System Identification and Pre-Distortion in Cognitive Transmitters*”. Journal of Lightwave Technology, 39(15), 5008–5020, 2021.

becomes a physically constrained excitation for the electro-optic interface (e.g., the modulator electrodes). Conceptually, this block reflects the deterministic transformation imposed by the electrical front-end, such as linear amplification and limited bandwidth, before any optical transfer function is applied. In high-baudrate PAM operation, this stage is not a minor implementation detail: its gain, frequency response, and group-delay characteristics directly influence the time-domain waveform, the available spectrum for modulation, and the amount of intersymbol interference (ISI) seen at the modulator input.

Accordingly, the driver is modeled as an analog front-end with a linear path that can be described by a gain and an effective LTI transfer function, with optional extensions for nonlinear compression/clipping and additive electrical noise in later sections. This separation is intentional: it allows the thesis to (i) treat bandwidth limitation as a “channel-like” mechanism that introduces memory and ISI in the baseband waveform, while (ii) keeping nonlinear and stochastic effects as separate factors that can be enabled, calibrated, and optimized independently when creating ML training data or assessing design trade-offs across the transmitter stack.

5.4.1 Linear gain and bandwidth limitation (filtering/ISI viewpoint)

Linear driver abstraction as an LTI front-end

Let $x_{\text{DAC}}(t)$ denote the continuous-time electrical waveform at the DAC output (i.e., after the DAC’s sample/hold and its internal reconstruction modeling). The linear part of the driver can be expressed as a gain G_{drv} followed by an effective analog transfer function $H_{\text{drv}}(f)$ that aggregates the driver’s small-signal bandwidth limitation and any intended linear shaping:

$$v_{\text{out}}(t) = G_{\text{drv}}(x_{\text{DAC}} * h_{\text{drv}})(t), \quad V_{\text{out}}(f) = G_{\text{drv}}X_{\text{DAC}}(f)H_{\text{drv}}(f),$$

where $h_{\text{drv}}(t)$ is the impulse response associated with $H_{\text{drv}}(f)$. In the time domain, the “filtering / ISI” viewpoint follows immediately: if the response length of $h_{\text{drv}}(t)$ is not negligible compared to the symbol interval T_s , then the driver behaves as a *memory channel* whose output depends on a neighborhood of symbols rather than only the current one.

In system simulations, this analog behavior is represented at the waveform sampling rate F_s via a discrete-time equivalent $h_{\text{drv}}[n]$, so that

$$v_{\text{out}}[n] = G_{\text{drv}} \sum_k h_{\text{drv}}[k] x_{\text{DAC}}[n - k].$$

This LTI abstraction is not a heuristic convenience: it corresponds to the standard practice of describing linear front-ends by their frequency response and translating that response into an impulse-response model for time-domain prediction and equalization design⁶⁶.

Bandwidth limitation as a source of ISI in PAM signaling

For multilevel PAM (including PAM-4), limited electrical bandwidth causes waveform smoothing and symbol-to-symbol interference at sampling instants, which manifests as vertical eye closure and slope distortion during decision times. In practice, high-speed PAM transmitter research often attributes pre-FEC performance loss at high symbol rates to combined limitations of the DAC, driver, and modulator electrical interface, leading to the use of explicit linear equalization (pre- and/or post) to restore usable eye openings⁶⁷.

The surrogate's driver block is therefore placed immediately after the DAC module to ensure that (i) the DAC-generated waveform is subjected to a well-defined linear analog constraint before the electro-optic mapping, and (ii) any upstream DSP (including pre-emphasis or later DPD hooks) is evaluated against a consistent, physically interpretable front-end response rather than an idealized “flat” electrical path. This aligns with reported PAM-4 demonstrations in which

⁶⁶ J. Schoukens, K. Godfrey, & M. Schoukens, “Nonparametric Data-Driven Modeling of Linear Systems: Estimating the Frequency Response and Impulse Response Function.” *IEEE Control Systems*, 38, 49-88, 2018.

⁶⁷ X. Tang, Y. Qiao, Y.-W. Chen, Y. Lu, and G.-K. Chang, “Digital Pre- and Post-Equalization for C-Band 112-Gb/s PAM4 Short-Reach Transport Systems,” *Journal of Lightwave Technology*, vol. 38, no. 17, pp. 4683–4690, 2020.

severe bandwidth constraints require careful front-end equalization and an accurate representation of the analog response to achieve the target line rates⁶⁸.

Practical modeling choices for $H_{\text{drv}}(f)$ in a surrogate

Within a thesis-level surrogate, $H_{\text{drv}}(f)$ is typically implemented using low-order LTI prototypes that provide a controlled bandwidth and interpretable time-domain behavior. Common choices include:

- *Single-pole/multi-pole low-pass responses* (capturing dominant bandwidth limitation and roll-off),
- *Butterworth-type magnitude shaping* (maximally flat magnitude for a specified order),
- *Bessel-type phase/group-delay behavior* (when time-domain fidelity and reduced ringing are prioritized).

The essential point is not the specific prototype, but the explicit mapping from a small set of physically meaningful parameters (e.g., -3 dB frequency, pole locations, damping) to the resulting ISI/memory observed in the waveform. In measurement-driven workflows, the same abstraction supports fitting $H_{\text{drv}}(f)$ from step/frequency-response characterization and then using that fitted model for waveform-accurate prediction (including eye statistics)⁶⁹.

5.4.2 Nonlinear behaviors (saturation/clipping) and how they would be represented

Beyond the linear gain and bandwidth limitations discussed in 5.4.1, a practical post-DAC driver can exhibit large-signal nonlinear behavior that becomes performance-limiting at high symbol rates and large swing. The dominant manifestations are (i) gain compression (AM/AM distortion), (ii) rail-limited

⁶⁸ M. Xiang, S. Fu, O. Xu, J. Li, D. Peng, Z. Gao, Y. Wang, and Y. Qin, “Advanced DSP Enabled C-Band 112 Gbit/s/λ PAM-4 Transmissions With Severe Bandwidth-Constraint,” *Journal of Lightwave Technology*, vol. 40, no. 4, pp. 987-996, 2022.

⁶⁹ B. Shi, Y. Zhou, H. Sun, T. Nguyen, J.E. Schutt-Ainé, “Statistical Method for Eye Diagram Simulation in High-Speed Link Nonlinear System Applications,” *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 14, no. 11, pp. 2050-2061, 2024.

saturation/clipping, and (iii) nonlinear memory effects when the nonlinear output stage is combined with finite bandwidth, bias networks, and peaking/compensation networks. A key modeling goal of the surrogate is therefore not circuit-level fidelity, but a system-level behavioral representation that reproduces the main distortions observed in waveform metrics (eye closure), frequency metrics (spectral regrowth/IMD), and link metrics (EVM/BER).

Behavioral decomposition and placement in the surrogate

A compact, widely used abstraction decomposes the driver into a *linear dynamic path* (capturing gain and frequency response), followed by a *static or quasi-static nonlinearity* and an optional *hard limiter* that represents finite output swing. In discrete time, with the post-DAC waveform $x[n]$, a representative structure is

$$\begin{aligned} x_{\text{lin}}[n] &= g(x * h_{\text{drv}})[n], & x_{\text{nl}}[n] &= f(x_{\text{lin}}[n]), \\ x_{\text{out}}[n] &= \text{clip}(x_{\text{nl}}[n]; V_{\text{min}}, V_{\text{max}}) \end{aligned}$$

where $h_{\text{drv}}[n]$ models the driver's small-signal linear response and $f(\cdot)$ captures the large-signal AM/AM behavior. This “linear-then-nonlinear” ordering corresponds to a *Wiener-type* viewpoint when the nonlinearity is driven by the already-bandlimited waveform; the complementary *Hammerstein-type* ordering (nonlinear-then-linear) is also common in broadband transmitter behavioral modeling when the nonlinear device precedes dominant filtering⁷⁰.

Saturation and clipping models

In many electrical drivers, a substantial portion of the impairment can be explained by *finite swing* and gradual transition into compression. Two practical model families are:

⁷⁰ M. Isaksson, D. Wisell, and D. Rönnow, “A comparative analysis of behavioral models for RF power amplifiers,” IEEE Transactions on Microwave Theory and Techniques, vol. 54, no. 1, pp. 348–359, 2006.

(a) Hard clipping (rail limiter)

A direct representation of supply/stack limitations is the saturating limiter:

$$\text{clip}(u; V_{\min}, V_{\max}) = \begin{cases} V_{\min}, & u < V_{\min} \\ u, & V_{\min} \leq u \leq V_{\max} \\ V_{\max}, & u > V_{\max} \end{cases}$$

This captures abrupt clipping and is useful when the waveform occasionally exceeds the allowed swing due to peak-to-average power ratio (PAPR) or overly aggressive equalization.

(b) Smooth (“soft”) saturation

When compression is gradual, a differentiable soft-limiter better matches measured AM/AM. A convenient closed-form option is a rational soft saturation (one example):

$$f_{\text{soft}}(u) = \frac{u}{\sqrt{1 + \left(\frac{u}{V_{\text{sat}}}\right)^2}} \quad (\text{scaled by } V_{\text{sat}} \text{ as needed}),$$

or alternatively $f_{\text{soft}}(u) = V_{\text{sat}} \tanh(\alpha u)$. Such forms capture the transition into compression without creating an unrealistically sharp corner in the transfer curve.

Polynomial nonlinearity for moderate compression

When the driver operates mostly below rails but exhibits *moderate* gain compression, a low-order odd polynomial is a standard base model (odd symmetry is appropriate when the operating point yields approximately symmetric distortion around a common-mode reference):

$$f_{\text{poly}}(u) = a_1 u + a_3 u^3.$$

The a_3 term reproduces third-order distortion, which is typically the most relevant for in-band regrowth and IMD behavior at modest compression levels. A

large body of transmitter behavioral modeling literature uses polynomial-family representations as the simplest bridge from measured AM/AM to system-level impairment predictions⁷¹.

Nonlinear memory: memory polynomial/parallel Hammerstein form

At high symbol rates, the combination of nonlinearity and bandwidth limitation produces nonlinear ISI (memory) that is not captured by a purely memoryless $f(\cdot)$. A tractable and widely used model class is the memory polynomial / parallel Hammerstein representation, which keeps complexity manageable while capturing frequency-dependent nonlinear effects:

$$x_n[n] = \sum_{k=0}^{K-1} \sum_{\substack{p=1 \\ p \text{ odd}}}^P a_{p,k} x_{lin}[n-k] |x_{lin}[n-k]|^{p-1}$$

Here K is the memory depth and P the nonlinearity order. This structure is especially convenient because it aligns directly with common DPD parameterizations and can be fitted to measured data using least-squares-type procedures or learned variants⁷².

Relation to Volterra-type modeling and DSP/DPD hooks

The *Volterra series* is a more general framework for nonlinear systems with memory, often written (schematically) as

$$y[n] = \sum_k h_1[k] x[n-k] + \sum_{k_1, k_2, k_3} h_3[k_1, k_2, k_3] x[n-k_1] x[n-k_2] x[n-k_3] + \dots$$

⁷¹ J. Moon and B. Kim, “Enhanced Hammerstein Behavioral Model for Broadband Wireless Transmitters,” IEEE Transactions on Microwave Theory and Techniques, vol. 59, no. 4, pp. 924–933, 2011.

⁷² M. Isaksson, D. Wisell, and D. Rönnow, “A comparative analysis of behavioral models for RF power amplifiers,” IEEE Transactions on Microwave Theory and Techniques, vol. 54, no. 1, pp. 348–359, 2006.

and it can capture richer cross-terms than memory polynomials, at the cost of rapidly growing complexity. In short-reach PAM links, Volterra/Wiener-type nonlinear equalization is used to mitigate transmitter-induced nonlinear distortion and nonlinear ISI when operating near bandwidth/swing limits, which is why the surrogate maintains an explicit nonlinear-driver representation and a clear insertion point for nonlinear compensation (DPD and/or receiver equalization)⁷³.

5.4.3 Electrical noise sources and where they enter

After the DAC, the electrical driver/front-end is the first stage that conditions the continuous-time voltage presented to the electro-optic interface (e.g., the modulator driver input). In practice, this stage introduces additive electrical noise through a combination of (i) *passive thermal (Johnson–Nyquist) noise* associated with resistive elements and terminations, and (ii) *active-device noise* (output-buffer /amplifier noise), which can be represented as an output-referred voltage-noise PSD that is filtered by the driver’s own bandwidth limitation. In CMOS-oriented front-end modeling, these contributions are commonly expressed as sums of white (thermal) and low-frequency (1/f) terms, then mapped through the circuit transfer functions to obtain output-referred noise⁷⁴.

From noise PSD to a time-domain additive process

At the surrogate level, the most transparent abstraction is to treat the driver as an LTI (or weakly nonlinear) operator followed by an *additive voltage noise* term at the driver output:

$$v_{\text{out}}(t) = \left(v_{\text{sig}}(t) \star h_{\text{drv}}(t) \right) + v_n(t),$$

where $v_n(t)$ is modeled as a stationary random process characterized by an output-

⁷³ N. Stojanović, F. Karinou, Z. Qiang, and C. Prodaniuc, “*Volterra and Wiener Equalizers for Short-Reach 100G PAM-4 Applications*,” *Journal of Lightwave Technology*, 35, 4583-4594, 2017.

⁷⁴ A. Romanova and V. Barzdėnas, “*On Noise Modeling of Capacitive Feedback Transimpedance Amplifiers in CMOS*,” *Applied Sciences*, vol. 12, no. 19, art. 10186, 2022.

referred PSD $S_{v,\text{out}}(f)$. The key link from frequency-domain specifications to time-domain samples is the variance integral:

$$\sigma_v^2 = \int_{-\infty}^{+\infty} S_{v,\text{out}}(f) df \quad \text{or, more generally,} \quad \sigma_v^2 = \int_{-\infty}^{+\infty} |H_{\text{sh}}(f)|^2 S_{v,\text{src}}(f) df,$$

where $H_{\text{sh}}(f)$ captures the shaping imposed by the driver/output network (and any explicit reconstruction filtering). This method of obtaining variance from PSD via $|H(f)|^2$ integration procedure is standard in practical front-end noise calculations. It is exactly what enables the generation of a consistent *time-domain Gaussian noise sequence* $v_n[k] \sim \mathcal{N}(0, \sigma_v^2)$ at the simulation sampling rate⁷⁵.

A particularly useful specialization is the *white-noise-density* case, where an output-noise density e_n (in $V/(\sqrt{\text{Hz}})$) is provided or assumed flat across the relevant band. Then

$$\sigma_v^2 \approx e_n^2 B_n, \quad B_n = \frac{1}{|H_{\text{drv}}(0)|^2} \int_0^{+\infty} |H_{\text{drv}}(f)|^2 df$$

with B_n the (single-sided) equivalent noise bandwidth of the driver chain. This form is especially convenient in a surrogate because it makes the noise level *co-vary automatically* with the modeled bandwidth limitation (stronger low-pass filtering $\Rightarrow$ smaller $B_n \Rightarrow$ smaller integrated noise variance, all else equal).

Using noise figure as a parameterization hook

When the driver is characterized by a *noise figure* (NF), the same additive-noise abstraction can be parameterized through the noise factor (F) and the stage gain. By definition,

$$F = \frac{\text{SNR}_{\text{in}}}{\text{SNR}_{\text{out}}}, \quad \text{NF}_{\text{dB}} = 10 \log_{10} (F),$$

so that a specified NF is interpreted as an SNR degradation that can be converted into an output

⁷⁵ A. Fava, F. Centurelli, P. Monsurró, and G. Scotti, “A Compact Low-Power Chopper Low Noise Amplifier for High Density Neural Front-Ends,” *Sensors*, vol. 25, no. 4, art. 1157, 2025.

referred noise level once the input reference noise and gain conventions are fixed, this provides a practical bridge from instrument-style specs (NF, gain) to the time-domain additive term used in a system simulator⁷⁶.

Performance interpretation for PAM and eye-diagram observables

In a PAM transmitter chain, additive driver noise is most directly visible as *vertical eye closure* and increased error probability after slicing, because it perturbs the decision variable without necessarily changing the deterministic ISI structure. Physics-based transmitter modeling studies that include intensity/electrical noise sources explicitly show how added noise degrades eye-diagram opening and thus the achievable sensitivity/BER for OOK/PAM-class signaling, making the eye diagram a natural validation observable for the driver-noise block⁷⁷.

Where does this enter in the present MATLAB driver implementation?

In the present surrogate, the driver noise is injected at the driver output voltage (i.e., after the modeled gain/bandwidth limitation and any optional nonlinear limiting), so that the noise is carried forward exactly as it would be in hardware: it directly perturbs the electrical waveform delivered to the modulator interface. The implementation supports three complementary parameterizations that map cleanly onto the theory above:

- **Direct RMS voltage noise:** user supplies σ_v and the simulator draws $v_n[k] \sim \mathcal{N}(0, \sigma_v^2)$.
- **Noise-density mode:** user supplies e_n ($V/(\sqrt{\text{Hz}})$), and the simulator computes $\sigma_v^2 = e_n^2 B_n$ using the driver's equivalent noise bandwidth.
- **Noise-figure mode:** user supplies NF, and the simulator converts it to a noise factor F , then computes an output noise variance consistent with the modeled gain and effective bandwidth.

⁷⁶ R. Noe, “Extension for ‘Consistent Optical and Electrical Noise Figure’,” *Journal of Lightwave Technology*, vol. 42, no. 10, pp. 3540–3545, 2024.

⁷⁷ A. Grabowski, J. Gustavsson, and A. Larsson, “Large-Signal Equivalent Circuit for Datacom VCSELs – Including Intensity Noise,” *Journal of Lightwave Technology*, vol. 40(22), pp. 7342–7350, 2022.

This structure keeps the driver-noise model *identifiable and calibratable* (via e_n , NF , or σ_v) while remaining consistent with the driver’s frequency response through B_n .

Extension hook: non-Gaussian electrical noise

While Gaussian additive noise is the default and often adequate for aggregate thermal/active-device noise, real electrical environments can exhibit bursty or impulsive components (e.g., interference-like events) that are not well captured by a single Gaussian variance. Classical non-Gaussian noise models (e.g., Class-A/Class-B families) provide a principled route to extend $v_n(t)$ beyond Gaussian assumptions when measurement evidence justifies it, without changing the injection point or the surrounding driver architecture⁷⁸.

5.4.4 Interconnect elements

(RF cables/connectors/measured loss hooks)

Although the current surrogate lacks an explicit interconnect block, the post-DAC electrical path can be extended systematically and calibration-ready by treating RF cables, connectors, launches, and PCB traces as a linear two-port network, whose behavior is characterized by measured S-parameters across the relevant frequency band. In this framework, the connection from “driver output” to “modulator input” is not merely an abstract wire, but a measurable, frequency-dependent channel that causes insertion loss, ripple, group-delay distortion, and reflections. These effects directly lead to pulse distortion and ISI when broadband PAM waveforms are used. A practical de-embedding process for fixtures and interconnect segments can be developed entirely within the two-port/S-matrix

⁷⁸ D. Middleton, “Non-Gaussian Noise Models in Signal Processing for Telecommunications: New Methods and Results for Class A and Class B Noise Models,” IEEE Transactions on Information Theory, vol. 45, no. 4, pp. 1129–1149, 1999.

domain, incorporating combined time–frequency methods to differentiate fixture effects from the device or channel of interest⁷⁹.

Two-port representation and measurable transfer function

Let the interconnect be represented as a two-port with a scattering matrix

$$S(f) = \begin{bmatrix} S_{11}(f) & S_{12}(f) \\ S_{21}(f) & S_{22}(f) \end{bmatrix},$$

measured using a vector network analyzer (VNA) with reference impedance Z_0 (typically 50Ω). For many transmitter interconnects designed to be approximately Z_0 -matched at both ends, the *forward transmission* $S_{21}(f)$ provides a first-order description of the magnitude/phase shaping imposed on the waveform (insertion loss and dispersion). In more general cases, e.g., when launches, packaging, or the driver/modulator input deviates from Z_0 , the conversion from $S(f)$ to an equivalent *voltage transfer function* $H(f)$ is done by translating S-parameters to a chain/ABCD representation under specified source/load impedances, enabling a consistent baseband LTI model suitable for convolutional simulation. A rigorous two-port de-embedding methodology that treats the measured structure as cascaded networks (fixture + DUT + fixture) and extracts a “DUT-only” response from measured S-matrices is provided in *thru-based* approaches, including validated “*thru-only*” methods for PCB/microstrip fixtures⁸⁰.

From measured S-parameters to time-domain filtering (impulse response)

Once an interconnect-only frequency response is available (either $S_{21}(f)$ under matched assumptions or a derived $H(f)$ under specified terminations), the surrogate can incorporate it as an LTI block through its *impulse response*

⁷⁹ W. Zeng, Y. Wang, and Z. Wang, “A de-embedding method based on combining time and frequency domains,” *Scientific Reports*, 15, 18637, 2025.

⁸⁰ R. Tong, Z. Dai, J. Olsson, I. Huynen, R. Ruber, and D. Dancila, “Taper Transmission Line Based Measurement—A Thru-Only De-Embedding Approach,” *IEEE Transactions on Microwave Theory and Techniques*, 70(9), 4199 - 4210, 2022.

$$h(t) = \mathcal{F}^{-1}\{H(f)\}, \quad y(t) = (x * h)(t),$$

where $x(t)$ is the driver output waveform and $y(t)$ is the waveform delivered to the modulator input. This converts measured insertion-loss and phase-slope characteristics into an *ISI-producing channel* in a physically interpretable way (e.g., precursor/postcursor structure in $h(t)$). Time-domain reflectometry concepts that explicitly relate frequency-domain S-parameter measurements to time-domain responses provide a practical foundation for this conversion and for diagnosing which discontinuities dominate the distortion⁸¹.

Calibration/de-embedding and uncertainty hooks

A key reason to model interconnects via measured S-parameters is that it allows the surrogate to align with laboratory calibration practice. In real measurements, “interconnect” often includes launches, connector repeatability, and PCB fixture segments whose variations introduce *de-embedding error* and uncertainty envelopes in $S_{21}(f)$. Statistical circuit-model approaches quantify how fixture variation perturbs the de-embedded response and therefore the predicted insertion loss and ripple effects that would otherwise be misattributed to the driver or DAC model⁸². At higher frequencies (tens to hundreds of GHz), measurement practice further motivates explicit calibration/de-embedding hooks (e.g., thru-reflect-line (TRL) / Line-Reflect-Reflect-Match (LRRM) families and pad/access-line removal), reinforcing that the surrogate’s interconnect block should accept

⁸¹ W. Zeng, Y. Wang, and Z. Wang, “A de-embedding method based on combining time and frequency domains,” *Scientific Reports*, 15, 18637, 2025.

⁸² Y. Liu, S. Yong, H. Gao, S. Hinaga, D. Padilla, D. Yanagawa, J. Drewniak, and V. Khilkevich, “S-Parameter De-Embedding Error Estimation Based on the Statistical Circuit Models of Fixtures,” *IEEE Transactions on Electromagnetic Compatibility*, 62, 1459-1467, 2020.

calibrated S-parameter data rather than idealized cable models when the goal is hardware-faithful prediction⁸³.

Implementation concept inside the surrogate (planned)

Conceptually, the interconnect option will be implemented as a modular, swappable block that:

1. **Loads** a measured Touchstone dataset (.s2p) describing the interconnect/fixture over frequency.
2. **Applies** optional de-embedding (when calibration structures are available) to produce an interconnect-only response consistent with the measurement plane⁸⁴.
3. **Converts** the result into an LTI filter $H(f)$ and $h(t)$, enforcing causality/passivity as needed (e.g., via smoothing or rational fitting), then filters the waveform via convolution.
4. **Tags metadata** with dataset IDs (cable type, length, connector set, calibration method, measurement date, bandwidth), enabling traceable scenario definition and later optimization.

This approach keeps the driver model “device-centric” while allowing realistic, measurement-anchored channel effects to be inserted when reproducing lab setups (e.g., long RF cables, multiple adapters, or test fixtures).

5.5 Optical carrier and modulation (laser + bias + MZM)

The electrical waveform generated by the transmitter DSP/DAC/driver chain must be translated into the optical domain through a physically meaningful interface. In the proposed digital surrogate, this interface is implemented by three

⁸³ Q. Le, K. Feng, and T. Kämpfe, “Analysis of Calibration and De-embedding for On-Wafer Characterization up to 170 GHz in 22-nm FDSOI CMOS Technology,” 2025 IEEE 24th Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems (SiRF), 83-85, 2025.

⁸⁴ R. Tong, Z. Dai, J. Olsson, I. Huynen, R. Ruber, and D. Dancila, “Taper Transmission Line Based Measurement—A Thru-Only De-Embedding Approach,” IEEE Transactions on Microwave Theory and Techniques, 70(9), 4199 - 4210, 2022.

interconnected sub-blocks: (i) a continuous-wave (CW) optical carrier model (laser), (ii) a bias-control abstraction that sets the Mach-Zehnder modulator (MZM) operating point, and (iii) an intensity-modulation transfer function that converts an applied drive voltage into optical output power (or, equivalently, into an optical field envelope). The goal is not simply to “attach a laser” to the electrical signal, but to maintain the correct physical invariants used by direct detection, most importantly, that the photodiode responds to *optical intensity* (square-law detection), and that the modulator bias determines whether the voltage-to-intensity mapping is locally linear, compressive, or expansive.

This section is written for the primary target of the surrogate, IM/DD links employing multi-level amplitude modulation (e.g., PAM4), while remaining compatible with higher-order PAM and with the experimental bipolar PAM (BPAM) detection methodology used in this thesis. In BPAM-oriented experiments, bias selection is particularly critical because it determines how the modulator transfer function allocates optical power to “positive” and “negative” electrical excursions, thereby affecting the recoverability of signed information after direct detection. Accordingly, the subsections that follow establish: (1) the CW laser representation adopted in the surrogate (and why a low-pass equivalent model is both valid and necessary), (2) the bias-point definition and its interface to the modulator, (3) the MZM transfer function currently implemented, and (4) the mathematical forms of key MZM non-idealities to be included in the finalized surrogate.

5.5.1 CW laser model used in the surrogate (low-pass equivalent rationale)

Low-pass equivalent representation of an optical carrier

A physical optical carrier at center frequency f_0 (angular frequency $\omega_0 = 2\pi f_0$) oscillates far faster than any modulation bandwidth relevant to IM/DD systems. Directly simulating the passband field would require sampling at a rate commensurate with f_0 (hundreds of THz at 1550 nm), which is neither necessary nor computationally meaningful for system-level modeling. Instead, the surrogate adopts the standard *complex-envelope (low-pass equivalent)* representation, where the analytic optical field is expressed as

$$E_a(t) = \tilde{E}(t)e^{j\omega_0 t}, \quad E_{\text{phys}}(t) = \Re\{E_a(t)\},$$

and $\tilde{E}(t)$ is the slowly varying complex envelope containing the modulation. The validity condition is that the envelope bandwidth B is much smaller than the carrier frequency f_0 (i.e., $B \ll f_0$), which is always satisfied for IM/DD links. This representation is the natural starting point for analysis of fiber-optic communication systems because it preserves the modulation dynamics while eliminating the need to resolve the optical oscillations themselves⁸⁵.

Why square-law photodetection removes the optical carrier

The main reason the complex envelope suffices for IM/DD is due to the nature of photodetection. An ideal photodiode generates a current proportional to the instantaneous optical power, which is proportional to the squared magnitude of the incident optical field (considering responsivity and receiver electrical filtering). If the received analytic field is $E_a(t)$, the detected (unfiltered) photocurrent can be expressed as

$$i(t) = \mathcal{R} |E_a(t)|^2,$$

where $\mathcal{R}$ is the photodiode responsivity. Substituting $E_a(t) = \tilde{E}(t)e^{j\omega_0 t}$ yields

$$|E_a(t)|^2 = |\tilde{E}(t)|^2 |e^{j\omega_0 t}|^2 = |\tilde{E}(t)|^2.$$

Therefore, after the intrinsic electrical bandwidth limitation of the receiver (and any explicit low-pass filtering used in modeling), the detected signal depends on the *envelope intensity* $|\tilde{E}(t)|^2$ rather than on the carrier oscillation. This is precisely the quantity produced (or implied) by an intensity modulator driven by an electrical waveform in an IM/DD transmitter⁸⁶.

CW laser envelope model used in the surrogate

In the surrogate, the CW laser is modeled as an optical field envelope with constant mean power P_0 (in watts), represented in **field units** such that

⁸⁵ G. P. Agrawal, “*Fiber-Optic Communication Systems*”, 4th ed., Wiley, 2010.

⁸⁶ L. Chen, B. Krongold, and J. Evans, “*Performance Analysis for Optical OFDM Transmission in Short-Range IM/DD Systems*”, *Journal of Lightwave Technology*, vol. 30, no. 7, pp. 974–983, 2012.

$$\tilde{E}_{laser}(t) = \sqrt{P_0} e^{j\phi(t)} \Rightarrow |\tilde{E}_{laser}(t)|^2 = P_0.$$

This choice is deliberate: it ensures that subsequent optical blocks (e.g., an MZM transfer function) can operate naturally on either field amplitude or power, while remaining consistent with square-law detection at the receiver. In the current implementation, the laser block generates a sampled sequence $\tilde{E}_{laser}[n]$ at the global simulation sampling rate F_s , with $T_s = 1/F_s$ and $t_n = nT_s$. A “passband output” is optionally available for diagnostic visualization by forming $E_a(t) = \tilde{E}(t)e^{j\omega_0 t}$. Still, the system-level surrogate operates in the low-pass equivalent domain because all relevant IM/DD observables are preserved through $|\tilde{E}(t)|^2$ ⁸⁷.

Canonical phase-noise/linewidth model (Wiener phase)

Although ideal IM/DD with pure intensity modulation is insensitive to a common optical phase rotation (since $|\tilde{E}(t)e^{j\phi(t)}|^2 = |\tilde{E}(t)|^2$), the surrogate includes a phase-noise option because (i) it is required for coherent extensions of the framework, and (ii) phase fluctuations can convert into effective intensity noise in the presence of dispersive propagation, optical filtering, or modulator chirp in more complete models.

The standard communication-theoretic model represents laser phase noise as a *Wiener process*,

$$\phi(t + \Delta t) = \phi(t) + \Delta\phi, \quad \Delta\phi \sim \mathcal{N}(0, 2\pi\Delta\nu\Delta t),$$

where $\Delta\nu$ is the laser linewidth (Hz) and Δt is the time step. In discrete time with $\Delta t = T_s$, this becomes $\phi[n] = \phi[n-1] + \Delta\phi[n]$, with $\Delta\phi[n]$ i.i.d. Gaussian increments whose variance scales linearly with T_s . This model is consistent with the interpretation of a Lorentzian optical lineshape arising from random-walk phase

⁸⁷ G. P. Agrawal, “*Fiber-Optic Communication Systems*”, 4th ed., Wiley, 2010.

evolution, and it is widely used to connect time-domain phase statistics to linewidth and coherence measures⁸⁸.

Furthermore, the linewidth itself can be linked to the underlying frequency-noise statistics. A helpful perspective (especially for surrogate calibration) is that laser coherence can be equally described through the frequency-noise power spectral density and its connection to the optical lineshape/linewidth, allowing parameter identification from measured frequency-noise spectra rather than just from time-domain phase traces⁸⁹.

Canonical relative intensity noise model (RIN)

Relative intensity noise (RIN) captures stochastic fluctuations of the emitted optical power around its mean. A typical system-level model represents the instantaneous laser power as

$$P(t) = P_0(1 + \epsilon_{\text{RIN}}(t)), \quad \mathbb{E}[\epsilon_{\text{RIN}}(t)] = 0,$$

and defines the (single-sided) RIN power spectral density as the *normalized* power-noise spectrum (units 1/Hz, often reported in dB/Hz or dBc/Hz). Over an electrical noise-equivalent bandwidth (B), the corresponding power variance is obtained by integrating the RIN PSD,

$$\sigma_P^2 = P_0^2 \int_0^B S_{\text{RIN}}(f) df,$$

which explicitly demonstrates how the same laser can be “quiet” or “noisy” depending on the receiver bandwidth and the concentration of the RIN spectrum. In an envelope formulation consistent with square-law detection, RIN is naturally added as multiplicative amplitude (power) noise before modulation or, depending on the physical scenario, after modulation but before detection. While the current MATLAB implementation mainly focuses on the deterministic CW carrier (and

⁸⁸ L. Zhang, L. Chen, and X. Bao, “Unveiling Delay-Time-Resolved Phase Noise Statistics of Narrow-Linewidth Laser via Coherent Optical Time Domain Reflectometry”, *Optics Express*, vol. 28, no. 5, pp. 6719–6733, 2020.

⁸⁹ Z. Zhang and A. Yariv, “A General Relation Between Frequency Noise and Lineshape of Laser Light”, *IEEE Journal of Quantum Electronics*, early access, 56(3), 1100105, 2020.

optional phase noise), this RIN formulation provides the mathematical foundation for the final surrogate as optical noise modeling is expanded⁹⁰.

5.5.2 Bias control concept and bias-point definition (interface to the modulator)

In an intensity Mach–Zehnder modulator (MZM), the bias voltage sets the DC operating point of the interferometer, thereby determining the active region of the electro–optic transfer characteristic. In a system-level digital model, it is helpful to represent the modulator electrode voltage as a superposition.

$$V_{\text{MZM}}(t) = V_{\text{drive}}(t) + V_{\text{bias}}(t),$$

where $V_{\text{drive}}(t)$ is the information-bearing electrical waveform delivered by the DAC/driver chain, and $V_{\text{bias}}(t)$ is a (nominally) slowly varying control signal that sets the operating point and may include slow drift and fast perturbations. The role of $V_{\text{bias}}(t)$ is not merely “offsetting” the drive: it determines (i) the small-signal linearity region, (ii) the achievable extinction at the logical levels, and (iii) the sensitivity of the modulated waveform to bias fluctuations, hence it directly impacts eye opening and the effective transmitter nonlinearity experienced by the receiver⁹¹.

Bias-point definition. For the single-drive intensity MZM model used in this surrogate (transfer function detailed in Sec. 5.5.3), bias points are naturally defined with respect to the modulator half-wave voltage V_{π} through the phase shift they impose between the interferometer arms. In the voltage-domain representation adopted here (bias added inside the transfer argument), three canonical operating points are typically distinguished:

⁹⁰ V. Villenas, D. Wu, N. Gültekin, M. Riani, and A. Alvarado, “On Geometric Shaping for 400 Gbps IM-DD Links with Laser Intensity Noise”, Optical Fiber Communication Conference (OFC), Optica Publishing Group, paper Th1K.8, 2025.

⁹¹ Y. Fu, X. Zhang, B. Hraimel, T. Liu, and D. Shen, “Mach-Zehnder: A Review of Bias Control Techniques for Mach-Zehnder Modulators in Photonic Analog Links”, IEEE Microwave Magazine, vol. 14, no. 7, pp. 102–107, 2013.

- **Maximum-transmission (constructive) point:** $V_{\text{bias}} \approx 0$ (periodic equivalents exist), yielding maximum optical output for zero drive.
- **Quadrature point:** $V_{\text{bias}} \approx V_{\pi}/2$, yielding the maximum slope of the intensity transfer and therefore the most linear small-signal intensity modulation region around mid-transmission⁹².
- **Null (minimum-transmission, destructive) point:** $V_{\text{bias}} \approx V_{\pi}$, yielding minimum optical output for zero drive (again with periodic equivalents).

These points are not merely “settings”; they correspond to distinct transmitter regimes. Quadrature bias is the conventional choice for intensity PAM generation when one desires an approximately linear mapping between small voltage excursions and optical intensity changes. In contrast, *bipolar PAM (BPAM) with direct detection* requires a fundamentally different operating concept: to obtain an approximately *signed* (positive/negative) optical-field modulation around zero, the MZM is biased at the *null point* and driven with a small-signal voltage such that the output optical *amplitude* becomes approximately linear in the drive (via a sine-like expansion around the null)⁹³. This makes the bias point a first-order design parameter in BPAM, because any bias deviation from the null breaks the intended symmetry and effectively introduces an amplitude offset that reduces the eye opening after detection⁹⁴.

Bias control as a commandable block in the surrogate. From an engineering viewpoint, a surrogate intended to remain faithful under hardware non-idealities should treat bias generation as a **commandable control block** that outputs a bias setpoint in volts and can represent both: (i) the named operating points (max/quadrature/null), and (ii) a continuous “knob” that permits arbitrary $V_{\text{bias},0}$ selection. The latter is important because the *effective* operating point depends on the combined DC offsets and low-frequency distortions of the DAC and driver chain, as well as on the modulator’s own static imperfections (e.g., finite extinction

⁹² *Ibidem*.

⁹³ M. Secondini and E. Forestieri, “Direct Detection of Bipolar Pulse Amplitude Modulation”, *Journal of Lightwave Technology*, vol. 38, no. 21, pp. 5981–5990, 2020.

⁹⁴ *Ibidem*.

ratio and insertion loss handled in the transfer model). A practical surrogate, therefore, benefits from supporting bias parameterizations such as

$$V_{\text{bias},0} = V_{\text{null}} + \Delta V$$

or

$$V_{\text{bias},0} \text{ chosen to realize a target normalized transmission at } V_{\text{drive}} = 0,$$

so that bias can be swept/optimized under realistic transmitter constraints rather than being locked to a single idealized point⁹⁵.

Bias drift and fast perturbations. In real modulators, particularly LiNbO₃-based devices, bias drift arises from slow physical processes (often effectively modeled as a low-frequency dynamical process), making long-term stabilization and re-centering necessary for consistent modulation⁹⁶. In the surrogate, it is therefore appropriate to represent the applied bias as a stochastic process around the setpoint, for example

$$V_{\text{bias}}[n] = V_{\text{bias},0} + \Delta V_{\text{drift}}[n] + \Delta V_j[n],$$

where $\Delta V_{\text{drift}}[n]$ captures slow drift (e.g., a random walk $\Delta V_{\text{drift}}[n] = \Delta V_{\text{drift}}[n-1] + \sigma_{\text{dr}} w[n]$) and $\Delta V_j[n]$ represents faster “bias jitter” (often modeled as zero-mean Gaussian noise, optionally shaped by a low-pass to reflect actuator/loop bandwidth)⁹⁷. This abstraction is particularly relevant for BPAM-at-null operation, because the transfer slope and symmetry around the null make the modulated signal highly sensitive to bias deviations⁹⁸.

⁹⁵ Y. Fu, X. Zhang, B. Hraimel, T. Liu, and D. Shen, “*Mach-Zehnder: A Review of Bias Control Techniques for Mach-Zehnder Modulators in Photonic Analog Links*”, IEEE Microwave Magazine, vol. 14, no. 7, pp. 102–107, 2013.

⁹⁶ J. Teng, S. Wang, Z. Yin, W. Chen, G. Fan-Yuan, G. Guo, Z. Han, and G. Guo, “*Arbitrary bias control of LiNbO₃ based Mach-Zehnder intensity modulators for QKD system*”, EPJ Quantum Technology, vol. 10, no. 33, 2023.

⁹⁷ *Ibidem*.

⁹⁸ M. Secondini and E. Forestieri, “*Direct Detection of Bipolar Pulse Amplitude Modulation*”, Journal of Lightwave Technology, vol. 38, no. 21, pp. 5981–5990, 2020.

Control-loop realism (optional dither/pilot modeling). Many practical bias controllers employ a small bias dither/pilot tone to extract an error signal and lock the operating point; however, this actuation can itself introduce a measurable performance penalty whose magnitude depends on modulation format and controller settings⁹⁹. A surrogate aimed at predicting eye-closure trends should therefore allow an optional bias-control overlay, such as

$$V_{\text{bias}}(t) = V_{\text{bias},0} + A_d \cos(2\pi f_d t) + \Delta V_{\text{drift}}(t) + \Delta V_j(t),$$

with (A_d, f_d) and drift/jitter parameters chosen consistently with the intended control strategy and bandwidth.⁴ Finally, emerging device/platform solutions explicitly target *bias-drift mitigation* at the device level, reinforcing that DC bias stability is a primary limitation and not a secondary implementation detail¹⁰⁰.

5.5.3 MZM transfer function currently implemented

In the surrogate, the Mach–Zehnder modulator (MZM) is modeled as a *memoryless* electro–optic intensity modulator whose main nonlinearity comes from the interferometric recombination of two optical paths with a controllable phase difference. In its standard form, the input optical field is split into two parts, each gaining an arm phase, and then recombined. This results in an output intensity that depends sinusoidally on the differential phase between the two arms¹⁰¹. This static transfer function models the fundamental mapping from the core field to intensity, which an IM/DD receiver ultimately detects via square-law photodetection (Section

⁹⁹ J. Wu, L. Zhong, X. Dai, Q. Yang, M. Cheng, L. Deng, and D. Liu, “*Numerical Analysis of Dither Induced Penalty for Automatic Bias Control Methods*”, IEEE Photonics Journal, vol. 16, no. 1, 7100309, 2024.

¹⁰⁰ S. Sun, M. He, M. Xu, S. Gao, Z. Chen, X. Zhang, Z. Ruan, X. Wu, L. Zhou, L. Liu, C. Lu, C. Guo, L. Liu, S. Yu, and X. Cai, “*Bias-drift-free Mach–Zehnder modulators based on a heterogeneous silicon and lithium niobate platform*”, Photonics Research, vol. 8, no. 12, p. 1958-1963, 2020.

¹⁰¹ S. Vervoort, Y. Saalberg, and M. Wolff, “*Mach–Zehnder Modulator Output in Time and Frequency Domain—Calculation and Experimental Confirmation*,” Photonics, vol. 10, no. 3, Art. 337, 2023.

5.5.1), while differentiating the deterministic interferometric nonlinearity from bandwidth/memory effects (discussed in Section 5.5.4)¹⁰².

Let P_{in} denote the continuous-wave optical power at the MZM input (provided by the laser block) and let $V(t)$ denote the electrical drive waveform delivered by the driver stage. In an ideal balanced MZM, the normalized optical power transmission can be expressed as a cosine function of the interferometer phase difference; an equivalent and widely used representation is a squared-cosine mapping:¹⁰³

$$T_{\text{ideal}}(V(t)) = \cos^2\left(\frac{\pi V(t) + V_{\text{bias}}}{V_{\pi}}\right) \quad P_{\text{out,ideal}}(t) = P_{\text{in}} T_{\text{ideal}}(V(t)).$$

Here, V_{π} is the *half-wave voltage*, defined as the voltage that produces a π phase shift (hence switching the interferometer from constructive to destructive interference). The bias term V_{bias} sets the operating point on the sinusoidal transfer (e.g., maximum transmission, quadrature, or null), consistent with the bias-control interface described in Section 5.5.2.³ For example, under the above parametrization, choosing $V_{\text{bias}} = 0$, $(V_{\pi}/2)$, and V_{π} moves the operating point to maximum, quadrature, and null, respectively, while retaining the same underlying transfer nonlinearity.

Because the surrogate is currently focused on IM/DD operation, the modulator output is represented in *low-pass equivalent* form as a real, nonnegative optical *field envelope amplitude* proportional to the square root of power. This is a deliberate system-level abstraction: the optical carrier oscillation and absolute optical phase are not propagated through this block because, for direct detection, the dominant

¹⁰² A. Tibaldi, M. Ghomashi, F. Bertazzi, and G. Ghione, “Organic Electro-Optic Mach–Zehnder Modulators: From Physics-Based to System-Level Modeling,” *physica status solidi (a)*, vol. 218, art. 2100390, 2021, doi: 10.1002/pssa.202100390.

¹⁰³ G. P. Agrawal, “*Fiber-Optic Communication Systems*”, 4th ed., Wiley, 2010.

observable is the instantaneous optical power (and its baseband fluctuations)¹⁰⁴. Concretely, the block outputs

$$a_{\text{out}}(t) = \sqrt{P_{\text{out}}(t)},$$

which can be interpreted as the magnitude of the complex envelope of the optical field at the modulator output in the IM/DD-oriented surrogate interface.

Inclusion of insertion loss and finite extinction ratio (as implemented)

Real MZMs exhibit non-ideal on/off contrast and optical loss, so the surrogate augments the ideal interferometric law with two system-facing parameters commonly reported in device characterization: insertion loss (IL) and extinction ratio (ER). Practical MZM assessments routinely quantify these figures as they directly affect the available optical power budget and the achievable modulation depth¹⁰⁵.

In the current implementation, insertion loss is applied as a multiplicative attenuation in the power domain:

$$P_{\text{out}}(t) \leftarrow P_{\text{out}}(t) \cdot 10^{-\text{IL}_{\text{dB}}/10},$$

and finite extinction ratio is enforced by introducing a residual power floor at the null. Defining

$$\rho \triangleq 10^{-\text{ER}_{\text{dB}}/10} = \frac{P_{\text{min}}}{P_{\text{max}}},$$

the implemented normalized power scaling becomes

$$T(V) = (1 - \rho)T_{\text{ideal}}(V) + \rho, \quad P_{\text{out}}(t) = P_{\text{in}} \cdot 10^{-\text{IL}_{\text{dB}}/10} \cdot T(V(t)).$$

This form preserves the canonical squared-cosine transfer while ensuring that: (i) the maximum transmission remains normalized to unity (before IL), and (ii) the minimum transmission is limited to ρ , consistent with an ER-limited “off” state. In

¹⁰⁴ A. Tibaldi, M. Ghomashi, F. Bertazzi, and G. Ghione, “Organic Electro-Optic Mach–Zehnder Modulators: From Physics-Based to System-Level Modeling,” *physica status solidi (a)*, vol. 218, art. 2100390, 2021, doi: 10.1002/pssa.202100390.

¹⁰⁵ A. Tibaldi, M. Ghomashi, F. Bertazzi, and G. Ghione, “Plasmonic-organic hybrid electro/optic Mach–Zehnder modulators: from waveguide to multiphysics modal-FDTD modeling,” *Optics Express*, vol. 28, no. 20, pp. 29253–29271, 2020.

system simulations, these two parameters are often sufficient to capture first-order penalty trends in IM/DD link budgets and eye opening when the goal is to represent the dominant static nonlinearity and power scaling without yet invoking higher-order physical imperfections¹⁰⁶.

5.5.4 MZM non-idealities considered for the final version

The baseline MZM intensity-modulation mapping introduced in Section 5.5.3 is intentionally compact because it is sufficient for establishing the end-to-end IM/DD chain and for validating transmitter DSP blocks. However, in high-speed multilevel PAM links, the dominant performance penalties are often set not by the ideal $\cos^2(\cdot)$ law itself, but by *deviations* from the ideal interferometric balance and by *dynamic* electro-optic memory. Therefore, the final digital-surrogate version extends the modulator model by introducing a small set of physically interpretable parameters that capture the most impactful non-idealities reported in state-of-the-art IM/DD system studies and device analyses¹⁰⁷.

A convenient starting point is a field-interference representation with explicit arm transmissions:

$$E_{\text{out}}(t) = \frac{E_{\text{in}}(t)}{2} (\sqrt{\eta_1} e^{j\phi_1(t)} + \sqrt{\eta_2} e^{j\phi_2(t)}), \quad P_{\text{out}}(t) = |E_{\text{out}}(t)|^2.$$

In the ideal balanced case $\eta_1 = \eta_2$ and with a purely differential phase drive, this reduces to the standard intensity law $P_{\text{out}}(t) \propto \cos^2(\Delta\phi(t)/2)$. Non-idealities enter by (i) introducing imbalance in η_1, η_2 and/or in the drive applied to each arm, (ii) allowing the bias phase to deviate and drift around its target value, (iii) permitting an intensity-to-phase coupling (“chirp”) at the *field* level, and (iv) filtering the voltage-to-phase conversion through a finite electro-optic response¹⁰⁸.

¹⁰⁶ *Ibidem*.

¹⁰⁷ G. Rizzelli, P. Torres-Ferrera, F. Forghieri, and R. Gaudino, “An Analytical Model for Performance Estimation in Modern High-Capacity IMDD Systems”, *Journal of Lightwave Technology*, 42(5), 1443–1452, 2024.

¹⁰⁸ T. Murao, J. Ushida, H. Takahashi, M. Tokushima, A. Shiina, and T. Horikawa, “Understanding Inter-Arm Imbalance in Mach–Zehnder Silicon Optical Modulators: Impact on OMA and Frequency Chirp”, *Journal of Lightwave Technology*, 40(15), 5171–5189, 2022.

(1) Finite extinction ratio

A finite extinction ratio (ER) means that the “null” transmission is not truly zero, so the modulator exhibits a non-vanishing optical-power floor P_{min} . In system-level IM/DD models, this is commonly represented by a leakage term in the intensity transfer:

$$P_{out}(t) = P_{min} + (P_{max} - P_{min}) \cos^2 \left(\frac{\Delta\phi(t)}{2} \right), \quad ER = \frac{P_{max}}{P_{min}}.$$

This parametrization is particularly important when the link is peak-power constrained (a typical situation for short-reach IM/DD), because ER directly couples the achievable modulation depth and the effective optical swing (OMA) to clipping and rate loss¹⁰⁹.

(2) Bias error and bias drift

Let the intended operating bias correspond to a target static phase ϕ_b (quadrature or null, depending on the modulation/detection strategy). A practical modulator experiences both *static bias error* and *time-varying drift*, which can be modeled by:

$$\phi_b(t) = \phi_b + \Delta\phi_{static} + \Delta\phi_{drift}(t).$$

For the drift term, a low-frequency stochastic process is appropriate (e.g., a random walk or a first-order autoregressive/Ornstein–Uhlenbeck process), because bias controllers are designed to track slow variations while remaining insensitive to the data modulation. In bias-control literature, the bias deviation $\Delta\phi$ is explicitly linked to measurable spectral components created by pilot/dither tones, enabling a model-based relation between drift and the resulting harmonic content at the detected output¹¹⁰.

¹⁰⁹ T. Wiegart, F. Da Ros, M.P. Yankov, F. Steiner, S. Gaiarin, and R. Wesel, “*Probabilistically Shaped 4-PAM for Short-Reach IM/DD Links With a Peak Power Constraint*”, Journal of Lightwave Technology, 39(2), 400–405, 2020.

¹¹⁰ J. Svarny and S. Chladek, “*Model-Based Bias Controller for a Mach-Zehnder Intensity Modulator*”, Journal of Lightwave Technology, 40(3), 720–727, 2022.

From an IM/DD-PAM viewpoint, bias deviation is not merely a DC offset: it reshapes the local slope of the $\cos^2(\cdot)$ mapping, causing unequal effective level spacing and introducing deterministic distortion that reduces vertical eye opening. In dispersion-uncompensated links, bias deviation can also interact with the frequency-domain transfer function, thereby altering fading notches and equalization requirements¹¹¹.

(3) Quadrature error as a distortion mechanism

Quadrature is often selected to maximize local linearity of the intensity mapping for multilevel amplitude modulation. Departures from exact quadrature are captured by writing the operating phase as

$$\phi_b = \frac{\pi}{2} + \varepsilon,$$

where ε is the quadrature error. A rigorous way to expose the resulting nonlinear distortion is to expand the power transfer around the operating point in a polynomial in the small-signal drive $v(t)$ (or equivalently in $\Delta\phi(t)$). For MZMs, such Taylor expansions yield closed-form expressions for harmonic and intermodulation products as explicit functions of the bias phase and modulation amplitude. This provides a direct analytical bridge between “quadrature error” and PAM eye distortion/EVM degradation, because even-order terms re-emerge when $\varepsilon \neq 0$ and the mapping becomes asymmetrically nonlinear around the operating point¹¹².

(4) Chirp parameter α : field-level IM–PM coupling

Although IM/DD receivers detect intensity, transmitter-induced phase modulation is still relevant: it can translate into amplitude distortion after dispersive

¹¹¹ Y. Zhu, X. Miao, Q. Wu, L. Yin, and W. Hu, “Imbalanced Mach-Zehnder Modulator for Fading Suppression in Dispersion-Uncompensated Direct Detection System”, *Electronics*, 10(22), 2866, 2021.

¹¹² A.M. Gutiérrez, A. Brimont, J. Herrera, M. Aamer, D. Thomson, F. Gardes, G. Reed, J. Fédéli, and P. Sanchis, “Analytical Model for Calculating the Nonlinear Distortion in Silicon-Based Electro-Optic Mach–Zehnder Modulators”, *Journal of Lightwave Technology*, 31(23), 3603–3613, 2013.

propagation, and it can also modify the effective frequency response experienced by multilevel PAM waveforms. A field-level description is therefore essential when chirp needs to be represented. In the general interferometric model, chirp appears when the modulator drive not only causes an intensity change but also produces a drive-dependent output phase.

$$E_{\text{out}}(t) = \sqrt{P_{\text{out}}(t)} e^{j\phi_{\text{out}}(t)}.$$

A widely used engineering definition of the (intensity) chirp parameter relates the induced phase dynamics to the induced intensity dynamics:

$$\alpha \triangleq \frac{2 \frac{d\phi_{\text{out}}(t)}{dt}}{\frac{d}{dt} \ln P_{\text{out}}(t)},$$

with appropriate interpretation in the small-signal regime. In silicon MZMs, α is not a fixed constant but can depend on inter-arm imbalance (phase/absorption non-uniformity) and on the bias point; “generalized” chirp-parameter expressions derived from arm-dependent complex modulation are therefore preferred over ad-hoc constants when targeting surrogate calibration against measured devices¹¹³.

When ER is finite, the field-level formulation naturally couples α and leakage: both effects arise from non-ideal arm responses and both impact detected amplitudes after propagation. Analyses that explicitly combine chirp and finite ER at the modulator model level provide reusable formulations for incorporating both impairments consistently in a surrogate¹¹⁴.

¹¹³ T. Murao, J. Ushida, H. Takahashi, M. Tokushima, A. Shiina, and T. Horikawa, “*Understanding Inter-Arm Imbalance in Mach–Zehnder Silicon Optical Modulators: Impact on OMA and Frequency Chirp*”, *Journal of Lightwave Technology*, 40(15), 5171–5189, 2022.

¹¹⁴ M. Ganjali and S. Hosseini, “*Effects of frequency chirping and finite extinction ratio of optical modulators in microwave photonic IFM receivers*”, *Optics Communications*, 452, 380–386, 2019.

(5) Arm imbalance and differential time skew

Inter-arm imbalance can originate from unequal losses (amplitude mismatch), unequal phase efficiencies, and differential electrical delays between arms. A compact and simulation-friendly representation is obtained by allowing:

- $\eta_1 \neq \eta_2$ (static amplitude imbalance),
- $\Delta\phi(t)$ to include different drive gains per arm (phase-efficiency mismatch),
- and a relative skew (τ) between the arm drives, e.g. $v_2(t) = v_1(t - \tau)$.

Such imbalance terms modify both the DC transfer (including ER and OMA) and the effective frequency response. Importantly, for IM/DD PAM in dispersive channels, they can shift or suppress dispersion-induced fading notches. This makes imbalance a first-class parameter in a surrogate used to analyze equalization needs, operating-point sensitivity, and robustness margins¹¹⁵.

(6) Finite electro-optic bandwidth and memory

At high symbol rates, the MZM cannot be treated as memoryless. A practical approach is to filter the voltage-to-phase conversion through an electro-optic impulse response $h_{EO}(t)$ (or frequency response $H_{EO}(f)$):

$$\Delta\phi(t) = \frac{\pi}{V_\pi} (v(t) * h_{EO}(t)), \quad \Leftrightarrow \quad \Delta\Phi(f) = \frac{\pi}{V_\pi} V(f) H_{EO}(f).$$

In traveling-wave modulators, $H_{EO}(f)$ is shaped by RF attenuation, velocity mismatch, and distributed electrode effects; time-domain large-signal models make this link explicit by deriving both small-signal frequency response and the resulting eye-diagram distortion under high-speed drive¹¹⁶.

For system-level studies, it is also common to represent the aggregate opto-electronic bandwidth by standard low-pass prototypes (e.g., Bessel) to quantify ISI

¹¹⁵ Y. Zhu, X. Miao, Q. Wu, L. Yin, and W. Hu, “Imbalanced Mach-Zehnder Modulator for Fading Suppression in Dispersion-Uncompensated Direct Detection System”, *Electronics*, 10(22), 2866, 2021.

¹¹⁶ H. Bahrami, H. Sepehrian, C. Park, L. Rusch, and W. Shi, “Time-Domain Large-Signal Modeling of Traveling-Wave Modulators on SOI”, *Journal of Lightwave Technology*, 34(11), 2812–2823, 2016.

regimes and to compare signaling formats under “light” versus “strong” bandwidth constraints¹¹⁷. Complementary to that, device/electrode-oriented studies show how to extract and use electrode frequency response as an RF filtering block inside digital simulations, directly connecting modulator design to waveform distortion and performance¹¹⁸.

Why these non-idealities matter for AI-based optimization in a digital surrogate

Including MZM non-idealities is not only about realism; it directly enhances the learnability and transferability of AI-based optimization performed on the surrogate. First, many of the effects listed above (bias error, ER leakage, chirp, imbalance, bandwidth memory) create structured distortions that are parametric and therefore identifiable from measured waveforms; a surrogate that reveals these parameters can be calibrated to hardware, reducing the simulation-to-experiment gap that otherwise hampers data-driven optimization¹¹⁹.

Second, much of the modulator impairment landscape is driven by the bias-dependent nonlinearity of the MZI transfer function, which allows for systematic polynomial or harmonic descriptions around an operating point. This provides an analytically grounded method for building differentiable impairment blocks and generating training data across different operating conditions, both of which are

¹¹⁷ R. Gutiérrez-Castrejón, M. Saber, M. Alam, Z. Xing, E. El-Fiky, D. Ceballos-Herrera, F. Cavaliere, G. Vall-Llosera, L. Giorgi, S. Lessard, R. Brunner, and D. Plant, “*Systematic Performance Comparison of (Duobinary)-PAM-2,4 Signaling under Light and Strong Opto-Electronic Bandwidth Conditions*”, *Photonics*,8(3),81, 2021.

¹¹⁸ T. Gaur, P. Mishra, G. Hegde, and T. Srinivas, “*Modeling and Analysis of Device Orientation, Analog and Digital Performance of Electrode Design for High Speed Electro-Optic Modulator*”, *Photonics*,10(3), 301, 2023.

¹¹⁹ G. Rizzelli, P. Torres-Ferrera, F. Forghieri, and R. Gaudino, “*An Analytical Model for Performance Estimation in Modern High-Capacity IMDD Systems*”, *Journal of Lightwave Technology*, 42(5), 1443–1452, 2024.

essential when optimization relies on gradient-based learning or end-to-end differentiable models¹²⁰.

Finally, when the surrogate is used to optimize transmitter-side DSP (e.g., predistortion, level shaping, bias tuning, or joint setting of DAC/driver/modulator operating points), the “optimal” solution depends on the *specific* modulator nonlinearity and memory. Recent work treating modulator transfer nonlinearities as a key computational element (e.g., in photonic neural-network contexts) emphasizes that these nonlinearities are not merely nuisances; they are deterministic transformations that can be modeled, differentiated, and exploited, provided the model accounts for the correct bias dependence and the non-ideal transfer shape¹²¹.

5.6 Channel model (TX-side propagation and noise-injection strategy)

After optical modulation, the transmitted waveform is shaped by the propagation channel before reaching the receiver front end. In IM/DD systems, this channel is not only the transmission fiber itself, but also the aggregate of all *optical-path* elements between the modulator output and the photodetector input (patch cords, connectors/splices, couplers, optional inline amplifiers, and laboratory components used for controlled noise loading). The surrogate therefore treats the “channel” as a physically grounded block that maps the transmitted complex envelope to a received complex envelope while parameterizing the dominant fiber-induced effects, such as loss, chromatic dispersion, and (when relevant) Kerr nonlinearity, together with a noise-injection strategy that can be aligned with either (i) a target physical link length or (ii) a laboratory OSNR-loading practice. The

¹²⁰ A.M. Gutiérrez, A. Brimont, J. Herrera, M. Aamer, D. Thomson, F. Gardes, G. Reed, J. Fédéli, and P. Sanchis, “Analytical Model for Calculating the Nonlinear Distortion in Silicon-Based Electro-Optic Mach–Zehnder Modulators”, *Journal of Lightwave Technology*, 31(23), 3603–3613, 2013.

¹²¹ P. Kincaid, N. Andriolli, G. Contestabile, and L. De Marinis, “Addressing optical modulator nonlinearities for photonic neural networks”, *Communications Engineering*, 4,58, 2025.

underlying fiber-impairment taxonomy and the standard link-level abstractions used in system simulators follow classical fiber-optic communication theory¹²².

A key design principle is *calibratability*: the channel block is parameterized so that nominal coefficients (e.g., attenuation and lumped insertion losses) can be overridden by experimentally observed values, enabling measurement-informed operation and closing the gap between simulation and hardware. This is consistent with the broader “measurement-informed model/digital-twin” philosophy increasingly adopted in optical communications, where measured physical-layer quantities are used to refine model parameters and improve predictive accuracy¹²³.

5.6.1 Attenuation and measured-loss hooks

Attenuation as distributed loss

Fiber attenuation is fundamentally a distributed power decay along distance (z). In its most common form, the optical power satisfies

$$P(z) = P(0)e^{-\alpha z},$$

where α is the power-attenuation coefficient in Np/m. Equivalently, the optical *field envelope* (the quantity propagated by most envelope-based simulators) obeys

$$A(z, t) = A(0, t)e^{-\frac{\alpha}{2}z},$$

since $P(z, t) = |A(z, t)|^2$. This “*square-root law*” is vital in a surrogate implementation, because when loss is applied to a complex envelope, the multiplicative factor must be $e^{-\alpha z/2}$ in the field domain to reproduce $e^{-\alpha z}$ in the power domain¹²⁴.

¹²² G. P. Agrawal, “*Fiber-Optic Communication Systems*”, Wiley, 2010.

¹²³ M. S. Faruk and S. J. Savory, “*Measurement Informed Models and Digital Twins for Optical Fiber Communication Systems*”, *Journal of Lightwave Technology*, vol. 42, no. 3, pp. 1016–1030, 2024.

¹²⁴ G. P. Agrawal, “*Fiber-Optic Communication Systems*”, Wiley, 2010.

In link engineering practice, loss is often specified in dB/km. Using α_{dB} in dB/km and fiber length L in km,

$$P(L)_{\text{dB}} = P(0)_{\text{dB}} - \alpha_{\text{dB}}L,$$

and the conversion to Np/m is

$$\alpha = \frac{\ln(10)}{10} \frac{\alpha_{\text{dB}}}{1000}.$$

This conversion is convenient because propagation operators are usually expressed in linear units, while configuration and laboratory specifications are generally given in dB.

Lumped insertion losses and link-budget aggregation

Beyond distributed fiber loss, practical Tx→Rx optical paths also incur lumped insertion losses, including connector/splice, coupler, and device packaging /interface losses. These are naturally summed in dB.

$$P_{\text{rx}}(\text{dBm}) = P_{\text{tx}}(\text{dBm}) - \alpha_{\text{dB}}L - \sum_k IL_k + \sum_m G_m ,$$

where IL_k are lumped losses (dB) and G_m are lumped gains (dB) such as amplifier gain. In an envelope-domain simulator, each lumped term corresponds to a multiplicative scaling of the field by $\sqrt{10^{-IL_k/10}}$ (for loss) or $\sqrt{10^{G_m/10}}$ (for gain), ensuring power consistency across blocks¹²⁵.

Measured-loss hooks as a calibration interface

In a laboratory setting, especially when exploring operating points like bPAM where biasing and modulation depth are closely linked to hardware non-idealities,

¹²⁵ *Ibidem*.

purely nominal loss coefficients are often inadequate. Slight variations in connector loss, fiber coupling, or component insertion loss can shift received power and modify the effective noise loading needed to achieve a specific operating OSNR/SNR target. For this reason, the surrogate channel is designed to support *measured-loss hooks*: parameters such as effective span loss, connector/coupler loss, or total fiber-to-fiber insertion loss can be input from measurements and used to override nominal values without altering the physical modeling structure.

This approach aligns with measurement-informed modeling: instead of treating the channel as a fixed, purely theoretical mapping, selected parameters are updated using observed data to enhance realism and predictive accuracy, a core concept in modern optical digital-twin methodologies¹²⁶.

5.6.2 Chromatic dispersion modeling choice

Chromatic dispersion (CD) is a deterministic *linear* propagation effect that introduces a frequency-dependent group delay, i.e., a quadratic phase response with respect to baseband frequency. In the surrogate, CD is modeled at the optical-field (complex-envelope) level, because this is the natural representation for propagation operators and it remains compatible with both IM/DD and future coherent extensions. Under the standard slowly-varying-envelope assumption, second-order dispersion is captured by the transfer function

$$H_{\text{CD}}(\omega) = \exp\left(-j\frac{\beta_2 L}{2}\omega^2\right), \quad \omega = 2\pi f,$$

so that the dispersed envelope is $\tilde{A}_{\text{out}}(f) = \tilde{A}_{\text{in}}(f)H_{\text{CD}}(2\pi f)$. The dispersion parameter used in fiber specifications, D (ps/nm/km), is related to β_2 through the standard conversion

$$\beta_2 = -\frac{\lambda^2}{2\pi c}D,$$

¹²⁶ M. S. Faruk and S. J. Savory, “Measurement Informed Models and Digital Twins for Optical Fiber Communication Systems”, *Journal of Lightwave Technology*, vol. 42, no. 3, pp. 1016–1030, 2024.

which enables a consistent mapping between tabulated fiber data and the envelope – domain operator used in simulation¹²⁷.

In IM/DD links, CD is especially important because it converts phase distortions accumulated during propagation into *intensity distortions* after square-law photodetection. For conventional double-sideband intensity modulation, this mechanism yields an effective electrical-frequency-selective response often referred to as *dispersion-induced power fading*, with nulls that depend on D , L , λ , and modulation frequency. This effect is not a “receiver artifact” but a direct consequence of field propagation combined with direct detection, and it is a principal reason why equalization requirements and achievable reach can change abruptly with baud rate and wavelength¹²⁸.

The surrogate’s modeling choice is therefore to treat CD as a *first-principles linear operator on the transmitted field*, rather than as an ad-hoc electrical-domain filter. This supports both faithful time-domain waveform evolution and analytical interpretation. In particular, CD-driven feasibility bounds for multilevel PAM can be expressed using compact dimensionless products of the form $(\beta_2 B^2 L)$, where B is the symbol rate (or an equivalent bandwidth measure). Such bounds are useful for defining *physically meaningful* operating regions (e.g., when a given PAM order becomes fundamentally reach-limited without compensation) and for separating “channel-imposed” penalties from transmitter/receiver implementation penalties¹²⁹.

In the current channel-prototype, CD is implemented using the second-order dispersion coefficient β_2 as the dominant linear memory mechanism in standard single-mode fiber propagation. This aligns with the intended first-stage goal of the digital surrogate: reproduce the principal deterministic ISI mechanism arising from fiber propagation at high baud rates, and provide a controllable parameterization

¹²⁷ G. P. Agrawal, “*Fiber-Optic Communication Systems*”, 5th edition, Wiley, 2021.

¹²⁸ G. J. Meslener, “*Chromatic dispersion induced distortion of modulated monochromatic light employing direct detection*”, IEEE Journal of Quantum Electronics, vol. 20, no. 10, pp. 1208–1216, 1984.

¹²⁹ E. Granot and G. Katz, “*PAM-N—Fundamental Limits in Chromatic Dispersive-Uncompensated Channels*”, Applied Sciences, vol. 11, no. 6, Art. 2542, 2021.

that can later be calibrated (e.g., by matching measured end-to-end frequency response or time-domain eye-closure trends).

For the final version, two extensions are planned when the operating conditions justify them: (i) *wavelength-dependent dispersion* and dispersion-slope support (relevant for wide wavelength excursions and for cross-band comparisons such as O-band vs C-band), and (ii) *higher-order linear terms* (e.g., β_3) and polarization effects when required by the scenario. In addition, for extremely wide optical spectra or mixed optical–RF scenarios, simplified CD models that explicitly account for ultrawide spectral content can be used as validated abstractions to reduce computational load while preserving dispersion-induced distortion trends¹³⁰.

5.6.3 Nonlinearities relevant to IM/DD regimes (implementation roadmap)

In an IM/DD link, the term *nonlinearity* can refer to two different (and often linked) mechanisms: (i) *propagation nonlinearities* originating from the optical fiber, and (ii) *effective nonlinear distortions* caused by the combination of chromatic dispersion and square-law photodetection. The digital surrogate uses an optical-field representation for propagation, so that both mechanisms can be modeled consistently: fiber propagation is depicted in the complex-envelope domain, while the receiver (photodiode) converts the field into an electrical intensity waveform via square-law detection. This is especially important in dispersive IM/DD systems, where dispersion turns phase variations into amplitude fluctuations and the photodetection process creates beating terms that can act as nonlinear ISI. The foundation of the model is the standard slowly-varying-envelope description of guided-wave propagation¹³¹.

A) Fiber Kerr nonlinearities: SPM / XPM / FWM (propagation domain)

The dominant fiber nonlinearity in standard single-mode fiber at C-band is the instantaneous Kerr effect, for which the refractive index depends on optical

¹³⁰ H. H. Elwan, J. Poette, and B. Cabon, “Simplified Chromatic Dispersion Model Applied to Ultrawide Optical Spectra for 60 GHz Radio-Over-Fiber Systems”, *Journal of Lightwave Technology*, vol. 37, no. 19, pp. 5115–5121, 2019.

¹³¹ G. P. Agrawal, “*Fiber-Optic Communication Systems*”, 5th ed., Wiley, 2021.

intensity ($n = n_0 + n_2|E|^2$). In a scalar single-polarization envelope model, the field envelope $A(z, t)$ evolving along distance z (with retarded time t) is governed by a nonlinear Schrödinger-type model that combines attenuation, chromatic dispersion, and a Kerr term:

$$\frac{\partial A(z, t)}{\partial z} = -\frac{\alpha}{2}A(z, t) - j\frac{\beta_2}{2}\frac{\partial^2 A(z, t)}{\partial t^2} + j\gamma|A(z, t)|^2A(z, t),$$

where α is the power-attenuation coefficient, β_2 is the group-velocity-dispersion coefficient, and γ is the nonlinear coefficient¹³².

For a *single channel*, the Kerr term primarily produces *self-phase modulation (SPM)*, often interpreted via the instantaneous nonlinear phase rotation

$$\phi_{\text{NL}}(t) \approx \gamma L_{\text{eff}}|A(t)|^2, \quad L_{\text{eff}} = \frac{1 - e^{-\alpha L}}{\alpha},$$

which becomes system-relevant when dispersion and filtering convert the induced phase modulation into intensity distortion at the detector. For *multi-channel (WDM)* operation, the same Kerr mechanism gives rise to *cross-phase modulation (XPM)* (phase perturbation of one channel by the intensity of others) and *four-wave mixing (FWM)* (generation of new frequency components by nonlinear mixing), which can be interpreted as nonlinear signal-signal interactions and signal-noise interactions in dispersive links¹³³.

The current channel prototype implements scalar split-step propagation with dispersion and attenuation in the linear operator, and a Kerr-induced phase rotation for SPM. XPM and FWM are structurally enabled by configuration flags but remain physically inactive in the current single-channel regime.

¹³² *Ibidem*.

¹³³ A. Ghazisaeidi, “A Theory of Nonlinear Interactions Between Signal and Amplified Spontaneous Emission Noise in Coherent Wavelength Division Multiplexed Systems”, *Journal of Lightwave Technology*, vol. 35(23), pp. 5150–5175, 2017.

As an upgrade, we considered extending the field model to multi-channel/WDM propagation so that XPM/FWM terms emerge from the total optical field, and (if required by scenario) introduce polarization-aware propagation (Manakov/Jones formalism) to capture polarization-dependent nonlinear mixing and PMD interactions.

B) “IM/DD-specific” nonlinear distortions: dispersion + square-law detection (effective channel nonlinearity)

Even when Kerr propagation effects are weak, dispersive IM/DD links exhibit nonlinear distortion mechanisms that are not present in coherent detection, because the receiver discards the optical carrier phase and measures intensity. In dispersive double-sideband PAM systems, square-law detection produces *signal–signal beating interference* (SSBI) and additional nonlinear power-series terms that depend on the channel's dispersive memory; these terms are a principal motivation for nonlinear compensation strategies (e.g., nonlinear precoding or nonlinear equalization) in practical IM/DD designs¹³⁴. This is also why many short-reach IM/DD studies separate *fiber Kerr nonlinearity* (often negligible under peak-power-limited operation) from *dispersion-induced detection nonlinearity*, and focus on the latter as the dominant impairment once bandwidth limits, CD, and detection are accounted for¹³⁵.

Implementation status in the surrogate (current): by propagating the complex optical envelope through dispersion and then applying square-law detection in the receiver chain, the surrogate is structurally capable of reproducing dispersion-induced detection nonlinearities (including SSBI-like behavior) as an emergent effect of the modeling chain, rather than as an *ad hoc* impairment term.

¹³⁴ H. Xin, K. Zhang, D. Kong, Q. Zhuge, Y. Fu, S. Jia, W. Hu, and H. Hu, “Nonlinear Tomlinson-Harashima precoding for direct-detected double sideband PAM-4 transmission without dispersion compensation”, *Optics Express*, vol. 27, no. 14, pp. 19156–19167, 2019.

¹³⁵ T. Wiegart, F. Da Ros, M. P. Yankov, F. Steiner, S. Gaiarin, and R. D. Wesel, “Probabilistically Shaped 4-PAM for Short-Reach IM/DD Links with a Peak Power Constraint”, *Journal of Lightwave Technology*, vol. 39, no. 2, pp. 400–405, 2021.

Planned upgrade path (effective IM/DD nonlinearities): (i) introduce reduced-order, communication-oriented nonlinear channel abstractions (e.g., polynomial/Volterra-type equivalents) for fast parameter sweeps and ML training sets, and (ii) integrate nonlinearity-aware DSP hooks aligned with recent IM/DD literature that explicitly characterizes nonlinear distortion mechanisms in such systems¹³⁶.

5.6.4 Noise injection modes (AWGN vs ASE-like injection mirroring lab practice)

A key design choice in the channel block of a digital surrogate involves *where* and *how* stochastic disturbances are introduced. This decision implicitly determines the physics assumptions and the statistical structure that the receiver algorithms perceive. In this surrogate, we identify two complementary noise-injection modes: (i) an *equivalent additive white Gaussian noise (AWGN)* model added at a specific electrical/digital observation point, and (ii) an *ASE-like optical-field injection* designed to mimic laboratory practices of OSNR-controlled noise loading and to preserve the characteristic beat-noise mechanisms of optically amplified IM/DD reception. This dual-mode approach is not redundant; it provides both a computationally efficient way to develop algorithms and a physics-consistent method to reproduce OSNR-driven experimental behavior while exposing the receiver to realistic, possibly signal-dependent noise statistics¹³⁷.

AWGN-mode (electrical/digital equivalent noise). In AWGN-mode, the surrogate injects a zero-mean Gaussian sequence at a specified point in the chain (e.g., after the channel filter, after photodetection, or after sampling), with variance chosen to achieve a target SNR at that point. The goal is that many link-level sensitivity trends and DSP comparisons can be effectively studied under an

¹³⁶ J. Li, X. Su, T. Ye, H. Nakashima, T. Hoshida, and Z. Tao, “Characterization of Nonlinear Distortion in Intensity Modulation and Direct Detection Systems”, *Journal of Lightwave Technology*, vol. 41(11), pp. 3513–3521, 2023.

¹³⁷ J. Witzens, J. Müller, and A. Moscoso-Mártir, “Modification of Level Dependent ASE-Signal Beat Noise by Optical and Electrical Filtering in Optically Preamplified Direct Detection Receivers”, *IEEE Photonics Journal*, vol. 10, no. 1, pp. 1–16, 2018.

equivalent AWGN abstraction, as long as the injection point is clearly indicated and the relationship between the target metric (SNR/EVM) and the injected variance is consistent with the sampling rate and normalization used. However, a key caveat is that “white when injected” does not mean “white when observed”: any subsequent *optical or electrical filtering* will shape the noise spectrum, creating *colored noise* at the decision point. Therefore, even when AWGN is used as the injected process, its actual impact must be considered together with the filtering or equalization chain that follows.¹³⁸

ASE-like mode (optical-field injection to reproduce OSNR loading). In ASE-like mode, noise is injected into the **optical complex envelope** as a stationary, circular complex Gaussian process $n_{\text{ASE}}(t)$ that adds to the signal field $E_s(t)$. This reflects the standard EDFA noise model, where ASE is described by a (locally) flat optical PSD over the signal bandwidth and is conveniently parameterized through amplifier gain and noise figure. In such models, the ASE PSD is calculated from a “well-known” gain-NF expression and then propagated through the link/filter chain, with filtering determining the final noise coloring¹³⁹.

For IM/DD, the key reason to prefer ASE-like injection when emulating OSNR-limited experiments is that square-law photodetection produces beat noise terms that are *not equivalent* to simple additive noise on intensity. If the received optical field is

$$E(t) = E_s(t) + n_{\text{ASE}}(t),$$

then the detected current (up to responsivity scaling) is proportional to

$$|E(t)|^2 = |E_s(t)|^2 + 2\Re\{E_s(t)n_{\text{ASE}}^*(t)\} + |n_{\text{ASE}}(t)|^2.$$

The middle term is the *signal-ASE beat noise*, whose variance scales with signal level, and the last term is the *ASE-ASE beat noise* contribution. Importantly, optical and electrical filtering can modify these contributions and produce counterintuitive changes in measured quality factors in practical receiver/test setups. This

¹³⁸ *Ibidem.*

¹³⁹ *Ibidem.*

mechanism is central to optically preamplified direct-detection receivers. It is a major reason an ASE-like option is necessary when the surrogate aims to reproduce OSNR-driven lab behavior rather than merely achieve an equivalent electrical SNR¹⁴⁰.

OSNR-controlled emulation of the lab noise-loading strategy. In our BPAM-oriented experiments, OSNR is swept by *noise loading with EDFA-generated ASE* and *power conditioning with variable optical attenuation*, with OSNR monitored by an OSA. A natural surrogate counterpart is therefore an OSNR-driven configuration in which the injected ASE level is set to meet a target OSNR defined in a reference bandwidth (as in standard OSNR practice). In this setting, the surrogate can expose two consistent “knobs” that match the experimental degrees of freedom: (i) amplifier/noise-figure-driven ASE generation (physics-parameter mode), and (ii) OSNR-targeted ASE scaling (measurement-metric mode). In both cases, preserving the ASE statistics in the optical field is valuable because it maintains the correct downstream beat-noise structure after photodetection and filtering. Additionally, when tight optical filtering or narrowband conditions are relevant, ASE may deviate from the simplest idealized assumptions (e.g., pulsing phenomena in narrow-band ASE), motivating careful interpretation of Gaussian approximations and bandwidth assumptions in surrogate-to-lab comparisons¹⁴¹.

Relevance to optimization and future extensions. Finally, supporting both AWGN and ASE-like injection aligns with the broader digital-surrogate philosophy: a surrogate designed for calibration, sensitivity analysis, and (future) data-driven optimization benefits from a configurable hierarchy of noise realism, ranging from equivalent AWGN for quick sweeps to amplifier/OSNR-driven ASE for accurate reproduction of physical-layer behavior and system-control scenarios. This aligns with the direction of modern physical-layer modeling used in open/disaggregated optical-network contexts, where ASE accumulation and

¹⁴⁰ *Ibidem*.

¹⁴¹ L. M. Muñoz-Cánovas, Y. O. Barmenkov, A. V. Kiryanov, J. L. Cruz, M.V. Andrés “*ASE narrow-band noise pulsing in erbium-doped fiber amplifier and its effect on self-phase modulation*”, *Optics Express*, vol. 27, no. 6, pp. 8520–8528, 2019.

parameterized amplifier behavior are fundamental components for performance prediction and optimization workflows¹⁴².

¹⁴² A. D’Amico, B. Correia, E. London, E. Virgillito, G. Borraccini, A. Napoli, and V. Curri, “*Scalable and Disaggregated GGN Approximation Applied to a C+L+S Optical Network*”, *Journal of Lightwave Technology*, vol. 40, no. 11, pp. 3499–3511, 2022.

Chapter 6

SOA Preamplifier, Receiver front-end, and Analog to Digital converter

Chapter 6 completes the receive-side part of the proposed digital surrogate by clearly modeling the electro-optical and mixed-signal blocks that lie between the optical waveform at the fiber output and the final digital samples used for detection. Instead of viewing the receiver as a perfect “measurement endpoint,” the surrogate captures the main physical mechanisms that alter statistics and degrade performance in high-speed IM/DD and multilevel PAM (including bipolar/unconventional variants): pre-amplifier gain dynamics and noise, photodetection and transimpedance conversion, bandwidth-induced memory, and the sampling/quantization limits imposed by the ADC. This level of realism is crucial for end-to-end optimization because the best operating point of each block (gain, bandwidth, clipping margin, effective resolution, etc.) is closely linked to the upstream waveform shaping and the downstream equalization/decision strategy.

The chapter is organized as follows. Section 6.1 introduces the SOA as a pre-amplifier and defines the gain, saturation, memory, and ASE-noise abstractions used in the surrogate. Section 6.2 summarizes the main trade-offs in operating regimes and their effects on eye/OSNR-related metrics. Section 6.3 develops the direct-detection electrical front-end, connecting the PD responsivity model to a TIA-based voltage output with physically grounded noise terms and bandwidth shaping, and relates these effects to eye/Q/BER behavior. Finally, Section 6.4 formalizes the ADC as an explicit sampling and quantization stage, including key non-idealities, so that the generated data stay hardware-consistent and suitable for ML training and transfer learning from simulation to laboratory measurements.

6.1 SOA Model in the Digital Surrogate

In the proposed IM/DD digital surrogate, the semiconductor optical amplifier (SOA) is used as a receiver-front-end component that can (i) increase the received optical power before square-law detection and (ii) simultaneously introduce *gain nonlinearity*, *data-dependent amplitude distortion*, and *amplified spontaneous emission* (ASE). These factors jointly determine the post-detection electrical SNR and the resulting BER/eye opening for multi-level signaling (PAM4 and BPAM). In access-oriented implementations (e.g., WDM-PON receivers), integrating the SOA also provides a practical way to balance the optical power budget while keeping the front end compact and scalable¹. Importantly for AI-driven optimization, SOA-induced impairments are not simply “noise-like”: at high symbol rates, the gain dynamics can create *pattern-dependent memory* that acts as data-dependent ISI, supporting surrogate models that reveal physically meaningful parameters (gain, saturation, noise factor, recovery time proxies) as adjustable factors for learning-based tuning and compensation².

6.1.1 SOA role in IM/DD receiver chains

In IM/DD links, the receiver primarily detects optical intensity after photodetection; therefore, any optical pre-amplification must be evaluated not only by its optical gain but also by how it alters the statistics of the detected current. In a pre-amplified direct-detection chain, the SOA is positioned before the photodiode (and any integrated optical filtering) to boost the optical power reaching the detector, with the main goal of enhancing receiver sensitivity (minimum received power at a target BER) and reducing the link loss budget. This architectural choice is especially important in access/fronthaul applications, where low-gain SOA

¹ P. Nadimi Goki, M. Imran, C. Porzi, V. Toccafondo, F. Fresi, F. Cavaliere, and L. Potì, “*Lossless WDM PON Photonic Integrated Receivers Including SOAs*”, Applied Sciences, vol. 9, no. 12, Art. 2457, 2019.

² L. Xue, L. Yi, R. Lin, L. Huang, and J. Chen, “*SOA pattern effect mitigation by neural network based pre-equalizer for 50G PON*”, Optics Express, vol. 29, no. 16, pp. 24714-24722, 2021.

blocks can serve as enablers for “loss-compensated” photonic nodes and compact receiver solutions³.

A system-level description begins by recognizing that the SOA acts as a *noisy, nonlinear optical gain element*. Denoting the input optical field as $E_{\text{in}}(t)$, a first-order system abstraction expresses the output as

$$E_{\text{out}}(t) \approx \sqrt{G}E_{\text{in}}(t) + n_{\text{ASE}}(t),$$

where G is the (possibly power- and time-dependent) gain and $n_{\text{ASE}}(t)$ is the complex ASE field added by the amplifier. After square-law photodetection with responsivity R , the electrical current becomes $i(t) = R|E_{\text{out}}(t)|^2$, which naturally produces *beating terms* between the signal and ASE as well as ASE-ASE beating. This is the key reason why “more gain” is not monotonic in benefit: the SOA can improve the mean detected signal swing while simultaneously increasing the variance through signal-ASE beat noise in a way that depends on the *optical and electrical filtering* present in the receiver⁴. For multilevel PAM (and especially for BPAM operating points that deliberately exploit symmetric electrical swings after detection), these filtering-dependent beat-noise effects translate into level-dependent noise and unequal eye openings across symbol levels. They must therefore be explicitly represented in a surrogate designed for BER-accurate data generation.

From a design perspective, the SOA’s role can be summarized as a three-way trade-off:

1. **Sensitivity vs OSNR trade-off.** Optical gain increases the detected signal amplitude, but ASE generated by the SOA degrades the effective OSNR at the detector input. After detection, the dominant impairment can

³ P. Nadimi Goki, M. Imran, F. Fresi, F. Cavaliere, and L. Potì, “Lossless ROADM by Exploiting low gain SOAs in fronthaul network”, in Proc. 2019 OptoElectronics and Communications Conference (OECC) / 2019 International Conference on Photonics in Switching and Computing (PSC), 2019.

⁴ J. Witzens, J. Müller, and A. Moscoso-Martir, “Modification of Level Dependent ASE-Signal Beat Noise by Optical and Electrical Filtering in Optically Preamplified Direct Detection Receivers”, IEEE Photonics Journal, vol. 10, no. 1, 7500416, 2018.

become signal-ASE beat noise, whose impact is strongly shaped by receiver filtering⁵.

2. Dynamic range and linearity vs gain. In practical operating ranges, the SOA gain compresses as input power increases, so the same element that helps close the power budget can also compress the PAM levels, causing deterministic eye closure and effective EVM penalties (developed in Secs. 6.1.3-6.1.4).

3. Access/PON relevance without deep device physics. Modern access-oriented receiver concepts explicitly consider SOA-based preamplification and integration to meet stringent split and reach requirements; this positions the SOA not as a laboratory-only “booster” but as a realistic candidate component in deployable IM/DD receiver chains⁶.

Within the digital surrogate, this subsection motivates why the SOA must be modeled in a *physics-informed but system-facing* manner: it must expose gain/ASE/linearity knobs that map onto measurable IM/DD performance metrics (eye opening, SNR, BER), and it must remain compatible with subsequent AI-based optimization, where learning algorithms benefit from structured, interpretable impairments rather than an undifferentiated noise lump⁷.

6.1.2 Gain model and carrier-dynamics abstraction

A semiconductor optical amplifier (SOA) is, in principle, a distributed traveling-wave device whose optical gain is governed by the carrier population in the active region and its interaction with the propagating photon flux. In full physical descriptions, the optical field evolution along the longitudinal coordinate z is coupled to a carrier-density rate equation $N(t)$, typically through a material gain term $g(N)$ and confinement factor Γ . A representative starting point (for a

⁵ *Ibidem*.

⁶ P. Nadimi Goki, M. Imran, C. Porzi, V. Toccafondo, F. Fresi, F. Cavaliere, and L. Potì, “Lossless WDM PON Photonic Integrated Receivers Including SOAs”, *Applied Sciences*, vol. 9, no. 12, Art. 2457, 2019.

⁷ L. Xue, L. Yi, R. Lin, L. Huang, and J. Chen, “SOA pattern effect mitigation by neural network based pre-equalizer for 50G PON”, *Optics Express*, vol. 29, no. 16, pp. 24714-24722, 2021.

single polarization and a slowly varying complex envelope $A(z, t)$) is the pair: (i) a propagation equation and (ii) a carrier equation:

$$(i). \quad \partial A / \partial z = 1/2 (\Gamma g(N) - \alpha_{\text{int}}),$$

$$(ii). \quad dN/dt = \frac{I}{qV} - \frac{N}{\tau_c} - \left(\frac{\Gamma g(N)}{\hbar \omega A_{\text{eff}}} \right) |A|^2,$$

where α_{int} is the internal loss coefficient and τ_c is an effective carrier lifetime. For system-level IM/DD simulation, however, the surrogate adopts a reduced-order *input-output* abstraction that preserves the physically meaningful control knobs (unsaturated gain, saturation level, and carrier recovery time) while avoiding a full longitudinal discretization. This reduction is consistent with state-space system models that collapse the SOA into a lumped dynamic element with gain compression, explicitly identifying the assumptions required for the lumped representation (e.g., spatially averaged carrier density and a tractable gain-compression law) and enabling direct mapping to communications metrics such as eye closure and sensitivity⁸.

Within this abstraction, the key static characteristic is the saturated gain curve $G(P)$, where G denotes *power gain* and P is the relevant instantaneous optical power (often P_{in} , P_{out} , or an average over the device). A widely used steady-state form expresses gain saturation via a gain-compression factor (or equivalently a saturation power), yielding reduced models such as

$$G_{\text{ss}} \approx \frac{G_0}{1 + \frac{P}{P_{\text{sat}}}} \quad \text{or} \quad G_{\text{ss}} \approx \frac{G_0}{1 + \epsilon P},$$

where G_0 is the small-signal (unsaturated) gain and P_{sat} (or ϵ) controls the onset/strength of compression. These compact expressions can be viewed as steady-

⁸ S. B. Kuntze, A. J. Zilkie, L. Pavel, and J. S. Aitchison, “Nonlinear State-Space Model of Semiconductor Optical Amplifiers With Gain Compression for System Design and Analysis”, *Journal of Lightwave Technology*, vol. 26, no. 14, pp. 2274-2281, 2008.

state consequences of the carrier-rate equation when the stimulated-emission term becomes comparable to the carrier injection/recovery terms, and are particularly attractive for a digital surrogate because they retain a clear parameter meaning and enable calibration to datasheet or measured gain-versus-power curves. Enhanced formulations explicitly incorporate *material-dependent* gain-compression behavior by augmenting the rate equations with an additional gain-compression term, improving agreement with experimental nonlinear SOA responses while remaining suitable for system simulation⁹.

In the implemented surrogate, the gain parameterization is organized around (i) an unsaturated gain reference $G_0(\lambda)$ and (ii) a saturation law expressed through P_{sat} , so that the SOA's static nonlinearity can be injected directly into the complex optical envelope used by subsequent blocks. Practically, $G_0(\lambda)$ serves as the primary “bias knob” (set by operating point and device type), while P_{sat} governs how rapidly amplitude statistics are compressed under large-signal conditions, which is an effect that translates into deterministic eye closure even in the absence of additive noise. This choice is consistent with experimentally validated static saturation modeling approaches that emphasize a system-relevant mapping from optical input/output powers to gain compression, enabling the surrogate to reproduce amplitude-dependent distortion with a small number of tunable parameters¹⁰.

Although the full carrier equation is not solved along z , the *carrier-dynamics content* of the physics is retained through a low-order dynamic extension of the static curve. In state-space terms, gain can be treated as a dynamical state that relaxes toward a power-dependent steady state, e.g.,

$$\frac{dG(t)}{dt} = \frac{G_{\text{ss}}(P(t)) - G(t)}{\tau_{\text{eff}}(t)},$$

⁹ O. M. Kharraz, A. S. M. Supa'at, A. Atieh, A.K. Zamzuri, and M. A. Mahdi, “Enhanced gain saturation model of non-linear semiconductor optical amplifiers”, IET Optoelectronics, vol. 12, no. 6, pp. 263-268, 2018.

¹⁰ C. Meuer, *et al.*, “Static gain saturation in quantum-dot semiconductor optical amplifiers”, Optics Express, vol. 16, no. 11, pp. 8269-8279, 2008.

where τ_{eff} represents an effective recovery constant capturing carrier depletion and recovery under modulation. In IM/DD, this abstraction bridges “*static gain saturation*” and “*pattern effects*”. When the symbol period approaches the recovery time, $G(t)$ cannot track G_{ss} instantaneously, creating data-dependent amplitude distortion that manifests as ISI-like memory. The present section establishes this modeling pathway (static curve + recoverable gain state) as the conceptual basis for the surrogate, while the explicit pattern-dependence mechanisms and their refinements are addressed in Section 6.1.4¹¹.

Finally, the gain abstraction in a complex-envelope surrogate naturally aligns with amplitude-phase coupling (through an effective linewidth-enhancement factor) and with noise-figure concepts that link gain to ASE generation. Even when the final receiver uses square-law (direct) detection, maintaining the complex envelope consistent with SOA gain/phase behavior is advantageous because it preserves the correct interface to dispersive optical propagation blocks and to OSNR-based noise-injection strategies. System-oriented analyses that jointly examine gain saturation and related performance metrics (including noise-figure trends and gain-dependent behavior) serve as the reference baseline for interpreting and calibrating these parameters when transitioning from a device description to a communication-system surrogate¹².

6.1.3 Saturation behavior and gain compression law

In IM/DD receiver chains, SOA gain saturation is a first-order system impairment because it couples *sensitivity extension* (more gain) to *nonlinear distortion* (gain compression), ultimately capping the usable receiver dynamic range in high-speed PAM links. In the linear regime, the SOA provides

¹¹ S. B. Kuntze, A. J. Zilkie, L. Pavel, and J. S. Aitchison, “Nonlinear State-Space Model of Semiconductor Optical Amplifiers With Gain Compression for System Design and Analysis”, *Journal of Lightwave Technology*, vol. 26, no. 14, pp. 2274-2281, 2008.

¹² J.-L. Xiao and Y.-Z. Huang, “Numerical analysis of gain saturation and noise figure in inhomogeneously broadened quantum-dot semiconductor optical amplifiers”, *IEEE Journal of Quantum Electronics*, vol. 44, no. 5, pp. 448 - 455, 2008.

approximately constant small-signal gain G_0 , so the amplified waveform maintains its multi-level amplitude spacing scaled by a factor. As the instantaneous optical power rises, stimulated emission depletes the carrier population, reducing the incremental gain available to the signal. The effect of multilevel intensity modulation is an amplitude-dependent transfer function: higher-power portions of the waveform experience lower effective gain than lower-power parts, leading to deterministic level compression and diminished eye opening at the upper levels. This distortion pattern most strongly penalizes PAM4 (and higher-order PAM) decision margins¹³.

A commonly used system-level abstraction is to parameterize the saturation using a gain compression law driven by either input or output optical power. In its simplest steady-state form, the gain is described as a decreasing function of optical power, for example:

$$G(P) = \frac{G_0}{1 + P/P_{\text{sat}}},$$

where P_{sat} is a saturation power that sets the onset and strength of gain compression. An equivalent representation is often given in terms of the 3-dB compression point $P_{3\text{dB}}$, defined as the input power at which the gain has dropped by 3 dB relative to G_0 , which is particularly convenient for mapping datasheet specifications into a simulation model. In practice, these reduced laws are used because they capture the dominant static nonlinearity with minimal parameter overhead, yet they remain consistent with the qualitative behavior observed in full carrier-rate-equation descriptions (linear amplification at low power, progressive compression at high power)¹⁴.

From a waveform-statistics viewpoint, gain compression reshapes the PAM amplitude distribution: the upper symbol levels and high-amplitude excursions are

¹³ Z. Li, Y. Li, S. Luo, F. Yin, Y. Wang, and Y. Song, “SOA Amplified 100 Gb/s/λ PAM-4 TDM-PON Supporting PR-30 Power Budget with >18 dB Dynamic Range”, *Micromachines*, vol. 13, no. 3, art. 342, 2022.

¹⁴ *Ibidem*.

“squeezed” more than the lower levels. This produces (i) reduced level spacing at the top of the eye, (ii) an effective dependence of optimal decision thresholds on operating point, and (iii) a distortion that is not well-represented by purely additive noise. System studies of pre-amplified PAM4 explicitly show that, beyond an optimal input-power region, performance degrades due to saturation-induced distortion, whereas the best operating region balances amplified signal power against nonlinear compression. This interplay is central to our surrogate because it determines whether the SOA should be modeled as a near-linear sensitivity booster (moderate gain, unsaturated) or as an intentionally “pushed” amplifier that trades eye linearity for power-budget extension¹⁵.

In long-reach and high-loss-budget IM/DD PON scenarios, the same mechanism becomes more essential because the SOA may need to operate closer to saturation to meet the receiver sensitivity target, especially when combined with aggressive symbol rates and multilevel modulation. Experimental demonstrations of SOA-enabled 100G-class PAM transmission show that the achievable power budget and receiver dynamic range are limited by SOA nonlinearities at higher output power, which motivates explicitly including saturation effects in the system model rather than treating the amplifier as a constant-gain block¹⁶.

In the digital surrogate, the saturation block is treated as a deterministic, parameterized nonlinearity that maps the incoming optical waveform power to an effective gain. The model is designed to be “physics-guided but system-level”: it captures the correct monotonic gain roll-off with power using a compact parameter set (G_0 plus a saturation/3-dB compression power), and it exposes these parameters as optimization variables. This is important for downstream AI-based optimization because the optimal operating point (gain setting/input power region) is not fixed; it depends on the upstream waveform statistics (PAM order, pulse shaping,

¹⁵ S. P. Ó Dúill, P. Landais, and L. P. Barry, “*Estimation of the Performance Improvement of Pre-Amplified PAM4 Systems When Using Multi-Section Semiconductor Optical Amplifiers*”, *Applied Sciences*, vol. 7, no. 9, art. 908, 2017.

¹⁶ A. G. Reza, L. N. Venkatasubramani, and L. P. Barry, “*112.5 Gbit/s long reach passive optical network with over 31 dB power budget enabled by semiconductor optical amplifiers*”, *Scientific Reports*, vol. 15, art. 27731, 2025.

clipping/pre-distortion decisions) and on how the receiver DSP/ADC chain tolerates nonlinear eye distortion. Once saturation is included, an optimizer can learn the true trade-space between gain, nonlinear distortion, and the noise terms that scale with optical power (treated in Sec. 6.1.5), instead of converging to unrealistically high-gain solutions that would be infeasible in a physical SOA¹⁷.

6.1.4 Pattern dependence and SOA “memory” effects

When an SOA is driven with intensity-modulated waveforms at symbol rates comparable to or faster than its effective carrier recovery dynamics, the gain cannot be considered an instantaneous function of the current input power. Instead, the amplifier shows *gain dynamics*, meaning that the instantaneous amplification depends on the *recent history* of the signal's optical power. This results in a *pattern-dependent response* in the form of identical symbols that can experience different gains depending on the preceding symbol sequence, and the output waveform inherits a data-dependent distortion that appears as nonlinear “memory” (i.e., a channel with state). In IM/DD links, especially multilevel formats such as PAM4 and experimentally extended bipolar/intensity-biased variants, this behavior can lead to level-dependent compression, unequal transition responses, and broadened spectra. These effects can tighten optical-filter constraints and cause measurable eye closure and sensitivity penalties¹⁸.

A widely used system-level abstraction is to represent the SOA gain by a *time-varying* gain state $G(t)$ (or logarithmic gain $g(t) = \ln G(t)$) whose evolution is governed by a first-order recovery term and a depletion term driven by the instantaneous optical power $P_{\text{in}}(t)$. In its simplest lumped form, this can be written as

¹⁷ V. V. Belykh, J. Rautert, V. S. Mikhlin, and A. E. Gubenko, “*Bit Error Rate in WDM Data Transmission Links With Semiconductor Optical Amplifier*”, Journal of Lightwave Technology, vol. 43, pp. 1076-1088, 2025.

¹⁸ Y. Fu, H. Xin, K. Zhang, J. Bai, M. Bi, X. Miao, L. Li, Z. Guo, H. He, and W. Hu, “*Impact of SOA-induced pattern effect on the filter requirements in vestigial sideband direct detected PAM4 transmission*”, Optics Express, vol. 26, no. 23, pp. 30305-30312, 2018.

$$\frac{dg(t)}{dt} = \frac{g_0 - g(t)}{\tau_c} - \kappa P_{\text{in}}(t), \quad G(t) = e^{g(t)},$$

where g_0 is the small-signal (unsaturated) gain, τ_c is an effective carrier-recovery time constant, and κ parameterizes gain depletion per unit input power. Equivalent formulations express depletion through a saturation energy/power (e.g., P_{sat}) and use a discrete-time recursion at the simulation sampling rate. The key systems insight is that memory becomes significant when τ_c is not negligible with respect to the symbol period T_s , because the gain state cannot fully reset between symbols; the resulting waveform distortion is therefore inherently sequence-dependent and behaves like nonlinear ISI. For PAM4, this creates asymmetric penalties across levels and transitions (e.g., more substantial distortion after high-power runs), whereas for BPAM-like operation (biasing near the null and exploiting sign through the modulator operating point), it can directly reduce effective eye opening by coupling pattern-dependent amplitude errors into the square-law detection process.

From a digital-surrogate perspective, capturing this phenomenon is not optional when the goal is *AI-based optimization* and data generation for realistic receivers. If the surrogate omits SOA memory, data-driven optimizers (predistortion, equalization, operating-point search, or joint Tx/Rx co-optimization) risk learning an *incorrect impairment map*, one that may look effective in simulation but fails to generalize when deployed at higher baud rates or different PRBS/payload statistics. Conversely, including pattern dependence enables the surrogate to generate training data that reflects the true nonlinear channel state, allowing ML models to exploit temporal context and learn mitigation strategies consistent with measured system behavior. This is consistent with recent experimental work demonstrating neural-network-based equalization/pre-equalization that explicitly targets SOA pattern effects in high-speed IM/DD PON scenarios¹⁹.

In the current surrogate design approach, the SOA block is structured to support both: (i) a memoryless regime, suitable for baseline OSNR/sensitivity studies and operation points far from dynamic limitations; and (ii) a gain-dynamics regime

¹⁹ L. Xue, L. Yi, R. Lin, L. Huang, and J. Chen, “SOA pattern effect mitigation by neural network based pre-equalizer for 50G PON”, Optics Express, vol. 29, no. 16, pp. 24714-24722, 2021.

where an internal gain state changes over time based on an effective recovery constant. This creates a clear path toward higher-fidelity upgrades, such as refined state equations with operating-point-dependent recovery, accounting for bandwidth-limiting interactions with post-SOA filtering, and calibration against measured patterning penalties, all while keeping the model manageable and suitable for systematic IM/DD optimization studies.

6.1.5 ASE noise generation in an SOA

In optically pre-amplified IM/DD receivers, amplified spontaneous emission (ASE) is not merely an additive power perturbation; it is a *stochastic optical field* that mixes with the modulated carrier through square-law photodetection and produces the dominant noise terms in many sensitivity-limited regimes. The surrogate therefore models ASE at the SOA output in a way that preserves (i) its *flat optical PSD* parameterized by gain and noise figure (or spontaneous-emission factor), and (ii) the *filter-consistent bandwidth integration* that determines the beat-noise statistics after direct detection. This is essential for producing calibration-faithful datasets and for training ML optimizers that must learn the correct trade-off between SOA operating point (gain/saturation) and receiver-limited noise penalties²⁰.

ASE PSD and its dependence on gain and noise figure / spontaneous-emission factor

A standard system-level representation treats ASE as a complex, band-limited Gaussian optical field with approximately white power spectral density across the optical noise bandwidth. In this formulation, the ASE variance (over an optical noise bandwidth B_o) can be written as

$$\sigma_{\text{ASE}}^2 = \frac{F}{2} (G - 1) h\nu B_o,$$

²⁰ P. Vidal-Moreno, C. Becerril, M. Fernández-Ruiz, H. Martins, S. Martín-López, and M. González-Herráez, “Noise analysis in direct detection and coherent detection phase-sensitive optical time-domain reflectometry systems.”, Optics Express, vol. 31, no. 17, pp. 27450-27461, 2023.

where G is the SOA power gain (linear), F is the *noise factor* (linear, with $NF = 10 \log_{10} F$ in dB), $h\nu$ is the photon energy at optical frequency ν , and the factor $1/2$ accounts for polarization conventions in the field model. This same modeling approach is equivalent to a flat ASE PSD proportional to $(G - 1)h\nu$, with the proportionality linked to F (or n_{sp}) depending on the chosen convention²¹.

For SOAs specifically, a more detailed perspective links ASE generation to the material gain and the spontaneous emission process, then relates the resulting ASE spectrum and power to a measurable system parameter (noise figure). This offers the theoretical foundation for viewing the SOA as a stochastic source whose ASE PSD and noise figure are aligned with its gain setting and carrier operating point, rather than as an arbitrary “added noise.” This connection is crucial in a digital surrogate: it limits the noise model so that ML-based optimization cannot leverage unphysical parameter combinations (e.g., very high gain with unrealistically low ASE)²².

Bandwidth integration consistent with optical filtering and receiver bandwidth

In a physically consistent IM/DD model, ASE is first integrated through the *optical filtering stage* (explicit or implicit) to obtain the in-band ASE field that actually reaches the photodiode. For an ideal rectangular optical filter of bandwidth B_{opt} , the in-band ASE power scales as

$$P_{ASE} \propto S_{ASE} B_{opt},$$

where S_{ASE} is the (approximately flat) ASE PSD. For non-rectangular optical

²¹ *Ibidem*.

²² A. De Melo and K. Petermann, “On the amplified spontaneous emission noise modeling of semiconductor optical amplifiers”, *Optics Communications*, vol. 281, pp. 4598-4605, 2008.

filters, the correct quantity is the *noise-equivalent bandwidth*, i.e., an integral of $|H_{\text{opt}}(f)|^2$ that captures how the filter weights the ASE spectrum²³.

After square-law detection, the photocurrent can be expanded as

$$i(t) = R|E_s(t) + E_{\text{ASE}}(t)|^2 = R|E_s(t)|^2 + 2R\Re\{E_s(t)E_{\text{ASE}}^*(t)\} + R|E_{\text{ASE}}(t)|^2,$$

which reveals two ASE – driven noise mechanisms:

- **Signal-ASE beat noise**, arising from the cross term $2R\Re\{E_s E_{\text{ASE}}^*\}$.

Its variance scales with signal level and depends on both optical filtering and the *electrical* receive filtering, typically through an electrical noise-equivalent bandwidth B_e . Importantly for PAM systems, this term becomes level-dependent, so the noise variance can differ across symbol levels (directly affecting optimal thresholds and eye closure)²⁴.

- **ASE-ASE beat noise** (spontaneous-spontaneous beating), arising from the $R|E_{\text{ASE}}|^2$ term. This contribution is strongly shaped by optical filtering and, in general, must be computed using filter-dependent noise-equivalent bandwidths; it can become dominant when the optical bandwidth is wide.²⁵

A key nuance for implementation is that optical filtering impacts these two contributions differently: tightening B_{opt} suppresses ASE-ASE beating more strongly than signal-ASE beating (because the former depends more strongly on the in-band ASE level). This is one reason why the surrogate must carry the *optical*-

²³ J.L. Rebola and A.V.T. Cartaxo, “*Q-factor estimation and impact of spontaneous-spontaneous beat noise on the performance of optically preamplified systems with arbitrary optical filtering*”, Journal of Lightwave Technology, vol. 21(1), pp. 87-95, 2003.

²⁴ J. Witzens, J. Müller, and A. Moscoso-Mártir, “*Modification of Level Dependent ASE-Signal Beat Noise by Optical and Electrical Filtering in Optically Preamplified Direct Detection Receivers*”, IEEE Photonics Journal, vol. 10, 7500416, 2017.

²⁵ J.L. Rebola and A.V.T. Cartaxo, “*Q-factor estimation and impact of spontaneous-spontaneous beat noise on the performance of optically preamplified systems with arbitrary optical filtering*”, Journal of Lightwave Technology, vol. 21(1), pp. 87-95, 2003.

filter bandwidth consistently through the chain rather than only matching a target OSNR at the end²⁶.

OSNR-referenced calibration and mapping to electrical SNR/BER

Because OSNR is the lab-facing quantity (typically measured by OSA in a reference bandwidth), the surrogate benefits from an explicit analytical mapping between OSNR and electrical SNR/BER for MZM-based IM/DD links, this provides a calibration hook: model parameters (G , F) (or (G, n_{sp})) can be adjusted such that the simulated OSNR and the predicted electrical SNR/BER follow the analytical dependence expected for the receiver configuration²⁷.

Interaction with channel AWGN: ordering and avoiding double-counting

When a separate AWGN channel noise option is used in the surrogate (e.g., as a controllable abstraction to match a target SNR/OSNR), it must be placed and interpreted consistently with the SOA ASE model to avoid double-counting.

- If AWGN represents propagation/launch-to-receiver noise accumulated before pre-amplification (e.g., an equivalent in-band optical noise floor reaching the SOA input), then it must be applied *before* the SOA, because it will be amplified and will interact with SOA saturation and filtering.
- If AWGN is used as an *equivalent receiver-input noise* calibrated to a measured OSNR/SNR at the photodetector (a common system-level abstraction), then it should be applied at the receiver input *after* the SOA stage, and the SOA ASE contribution must be disabled or reduced accordingly to preserve a one-to-one correspondence between physical noise sources and injected noise in the model.

²⁶ P. Vidal-Moreno, C. Becerril, M. Fernández-Ruiz, H. Martins, S. Martín-López, and M. González-Herráez, “Noise analysis in direct detection and coherent detection phase-sensitive optical time-domain reflectometry systems.”, *Optics Express*, vol. 31, no. 17, pp. 27450-27461, 2023.

²⁷ Y. Xie, S. Kandeepan, and K. Wang, “An Analytical Model of OSNR and SNR for Short Reach IM-DD Systems with MZM”, 2022 Asia Communications and Photonics Conference (ACP), pp. 1826-1830, 2022.

This ordering discipline is important for ML-based optimization. Otherwise, the optimizer may learn policies that exploit inconsistent noise accounting (e.g., “free” gain) and will not transfer to the laboratory setup where OSNR is controlled by amplifier ASE and optical filtering.

6.1.6 Polarization assumptions and practical implications

In this work, the receiver chain is modeled in single-polarization operation: the signal entering the SOA is considered to be polarized in a fixed state aligned with the model’s effective gain axis. This is a deliberate system-level abstraction that aligns with many IM/DD laboratory setups (short fiber runs, controlled patching, polarization-maintaining components, or stable polarization at the SOA input), and it lets the surrogate focus on the main IM/DD impairments, specifically gain dynamics, saturation, and ASE statistics, without mixing them with polarization-induced fading. However, polarization assumptions remain a *first-order design consideration* for a modular “device-equivalent” SOA block because the same SOA model might later be reused in (i) longer links with an uncontrolled state of polarization(SOP), (ii) WDM/PON deployments, or (iii) future dual-polarization architectures. Therefore, the surrogate must provide a clear interface and upgrade path for polarization-dependent behavior, even though this thesis’s scope remains limited to single-pol mode.

Polarization dependence in SOA vs surrogate simplifications

Physical SOAs generally exhibit *polarization-dependent gain (PDG)* because the guided modes experience different confinement factors and material gain for the transverse-electric (TE) and transverse-magnetic (TM) components, especially in strained quantum-well or strongly asymmetric waveguides. A compact system-level description models the SOA gain as polarization-selective with two principal axes (TE/TM). For an input field decomposed as

$$E_{\text{in}}(t) = E_{\text{TE}}(t)\hat{e}_{\text{TE}} + E_{\text{TM}}(t)\hat{e}_{\text{TM}},$$

a first-order polarization-dependent amplification can be written as

$$E_{\text{out}}(t) = \sqrt{G_{\text{TE}}} E_{\text{TE}}(t) \hat{e}_{\text{TE}} + \sqrt{G_{\text{TM}}} E_{\text{TM}}(t) \hat{e}_{\text{TM}},$$

where G_{TE} and G_{TM} are the (power) gains seen by the TE and TM components. A common PDG figure is then

$$PDG_{\text{dB}} = 10 \log_{10} \left(\frac{G_{\text{max}}}{G_{\text{min}}} \right) = 10 \log_{10} \left(\frac{\max(G_{\text{TE}}, G_{\text{TM}})}{\min(G_{\text{TE}}, G_{\text{TM}})} \right).$$

In IM/DD receivers, any SOP variation at the SOA input results in amplitude fluctuations at the SOA output, and, after photodetection, it becomes an effective *slow multiplicative distortion* of the signal intensity that can reduce eye opening and bias the estimated OSNR/SNR if it is not accounted for in the model. For this reason, even when the surrogate runs in single-pol mode, the SOA block is conceptually treated as a “re-usable device” whose polarization interface can later be activated by specifying $(G_{\text{TE}}, G_{\text{TM}})$ or a PDG parameter and a polarization state model²⁸.

In the current implementation scope, the surrogate simplifies the above by assuming a fixed SOP aligned to the effective gain axis, i.e., $G_{\text{eff}} \approx G_{\text{TE}}$ (or G_{TM}) and no stochastic SOP evolution. This is consistent with many silicon-photonics receiver prototypes and practical integrated SOA-based front-ends, in which polarization handling is either engineered (e.g., TE-optimized coupling) or constrained at the module level.

When polarization effects can be ignored vs when a penalty term is needed

Polarization effects can be neglected (without compromising model validity) when one of the following conditions holds:

1. **SOP is effectively fixed** over the observation time (e.g., in short patch-cord setups, PM fiber/components, or stable packaging, such that the TE/TM projection is constant). In this case, polarization dependence is

²⁸ A. Sobhanan, A. P. Anthur, S. Oduill, M. Pelusi, S. Namiki, L. Barry, D. Venkitesh, and G. P. Agrawal, “Semiconductor optical amplifiers: recent advances and applications”, *Advances in Optics and Photonics*, vol. 14, no. 3, pp. 571-651, 2022.

absorbed into the measured/fit gain and NF parameters used to calibrate the SOA block²⁹.

2. **PDG is intrinsically small** for the intended operating point (design or bias region where TE/TM gains are nearly equal), so that the residual gain variation across SOP is below the link margin and does not materially affect BER/eye metrics³⁰.

A penalty term (or an explicit SOP model) becomes necessary when:

- The SOP at the SOA input is *random or time-varying*, as in longer SMF runs with environmental perturbations. Then the effective gain becomes a random variable $G_{\text{eff}}(t)$, producing slow fading of the detected PAM levels and potentially shifting decision thresholds.
- The target scenario includes *reconfigurable use* of the SOA model as a “device-equivalent” block across different links and applications (e.g., changing fiber plant, different photonic front-ends, or dual-pol extensions). In such cases, a single-pol abstraction can be kept as a baseline mode, but the model should expose a switchable polarization layer that can generate G_{eff} either from TE/TM projections or from a PDG-parameterized statistical model³¹.

From a surrogate/AI perspective, representing polarization sensitivity is important for *dataset correctness and transferability*: if the training data never includes SOP-induced gain variability, an optimizer may learn operating points (e.g., aggressive gain/saturation settings) that look optimal in simulation but become fragile in real deployments where polarization fluctuations introduce additional amplitude wander and eye closure. Therefore, even while remaining in

²⁹ P. Nadimi Goki, M. Imran, C. Porzi, V. Toccafondo, F. Fresi, F. Cavaliere, and L. Potì, “*Lossless WDM PON Photonic Integrated Receivers Including SOAs*”, Applied Sciences, vol. 9, no. 12, 2457, 2019.

³⁰ A. Sobhanan, A. P. Anthur, S. Oduill, M. Pelusi, S. Namiki, L. Barry, D. Venkitesh, and G. P. Agrawal, “*Semiconductor optical amplifiers: recent advances and applications*”, Advances in Optics and Photonics, vol. 14, no. 3, pp. 571-651, 2022.

³¹ *Ibidem*.

single-pol mode in this thesis, the modular SOA block is structured so that polarization dependence can be introduced later as either (i) a physics-based TE/TM gain model with an SOP evolution process, or (ii) a controlled penalty abstraction parameterized by PDG and a SOP coherence time, without changing the surrounding Rx chain³².

6.1.7 Parameter choices and calibration hooks

A system-level SOA surrogate is only as reliable as the *parameter-to-phenomenon mapping* that ties its coefficients to measurable device behavior (gain, saturation, noise, bandwidth, and polarization sensitivity). In a “digital-equivalent device” philosophy, the SOA block must remain reusable across scenarios (e.g., different IM/DD link budgets, PON-like access setups, or receiver-front-end studies) without rewriting the model. This requires (i) a consistent mapping from datasheet parameters to model coefficients, and (ii) explicit *calibration hooks* that allow measurement-informed refinement so that the surrogate remains predictive when used as a dataset generator for AI optimization. The broader measurement-informed modeling perspective is well aligned with this approach: the surrogate is treated as a parametric physical model whose coefficients are updated using observed system behavior and device characterization, rather than as an unconstrained curve fit³³.

Mapping datasheet parameters to model coefficients (baseline parameterization)

For a first-principles-consistent system model, the baseline parameter set is chosen to correspond to standard SOA specifications:

- **Small-signal gain** G_0 : establishes the linear amplification regime and is the reference against which compression is defined.
- **Saturation metric** (e.g., P_{sat} or a compression point): determines when the SOA transitions from linear amplification to nonlinear gain

³² *Ibidem*.

³³ M. S. Faruk and S. Savory, “Measurement Informed Models and Digital Twins for Optical Fiber Communication Systems”, *Journal of Lightwave Technology*, vol. 42(3), pp. 1016-1030, 2024.

compression, which directly impacts PAM4/BPAM level spacing and eye opening.

- **Noise figure/spontaneous-emission parameter:** constrains ASE generation and prevents nonphysical “high gain, low noise” configurations that would corrupt OSNR-to-BER realism.
- **Effective bandwidth parameter:** captures the finite frequency response relevant for high-baud-rate IM/DD, where SOA dynamics and bandwidth interact with post-amplifier optical/electrical filtering.
- **Polarization sensitivity descriptors** (PDG/PDL, where applicable): indicate whether the gain/noise model should be interpreted as single-polarization-aligned or polarization-averaged.

This datasheet-to-model mapping is central for keeping the surrogate “physics-guided” while remaining implementable and calibratable at the system level, and it is consistent with methodology papers focusing on amplifier characterization/parameter identification for wideband applications³⁴.

Measurement-informed calibration: fitting gain, saturation, and noise behavior

In practice, datasheet values provide a starting point but do not fully capture the exact operating-point dependencies of a deployed device (bias current, temperature, coupling losses, and wavelength detuning). Therefore, the surrogate exposes *calibration degrees of freedom* via controlled scale/offset terms, enabling measured results to refine the effective parameters without changing the model structure. Typical calibration workflows include:

1. **Gain-curve fitting:** measure $G(P_{\text{in}})$ or $P_{\text{out}}(P_{\text{in}})$ at the relevant wavelength and bias point, and tune the surrogate’s effective G_0 and saturation metric so that the gain roll-off matches the measured curve.
2. **Noise alignment using OSNR/BER trends:** measure OSNR at the receiver input (or after optical filtering) and compare BER/eye-closure trends versus SOA settings; tune the surrogate noise parameter(s) (noise

³⁴ A. Beshr and M. Aly, “Characterization of wideband semiconductor optical amplifier: numerical analysis and simulation”, Optical and Quantum Electronics, vol. 55, 287, 2023.

figure / equivalent spontaneous-emission factor) so that OSNR-driven performance predictions match observed slopes and penalties.

3. **Bandwidth-consistent refinement:** use frequency-response or time-domain characterization to adjust the surrogate bandwidth parameter so the model reproduces high-speed penalties under realistic filtering and sampling conditions.

Recent work that explicitly combines SOA behavior with data-driven or AI-assisted modeling pipelines motivates this strategy: the surrogate remains parametric and interpretable, yet can be refined using measured responses to ensure that generated datasets reflect device-true operating regimes rather than merely nominal specifications³⁵.

EDFA-based pre-amplified setup as a calibration baseline and comparison point

Because the experimental BPAM/IM/DD setup uses an EDFA-based pre-amplification path (with ASE controlled via attenuation and OSNR monitored by OSA), the surrogate benefits from an **EDFA-calibrated baseline**: first match the measured OSNR-to-performance behavior under EDFA pre-amplification, and then transfer the same OSNR accounting into the SOA model by enforcing consistent bandwidth integration and noise-figure interpretation. This provides a controlled reference point for comparing “equivalent OSNR” performance with SOA pre-amplification, isolating the additional penalties that arise from SOA-specific nonlinearities and memory effects. The digital-twin viewpoint emphasizes exactly this kind of measurement-informed alignment: a model block is not validated only by local parameters, but by reproducing end-to-end trends under laboratory-relevant measurement definitions³⁶.

³⁵ R. Kraemer, O. Düzgöl, S. Li, and N. Calabretta, “Data-Driven SOA Parameter Discovery and Optimization Using Bayesian Machine Learning With a Parzen Estimator Surrogate”, *Journal of Lightwave Technology*, vol. 42, pp. 721-731, 2024.

³⁶ G. Borraccini, Y. Huang, A. D’Amico, T. De Lima, E. Ip, V. Curri, T. Wang, and K. Asahi, “Optical Line Physical Parameters Calibration in Presence of EDFA Total Power Monitors”, 2024 Optical Fiber Communications Conference and Exhibition (OFC), paper M31.5, 2024.

Why calibration hooks matter for dataset fidelity and AI-based optimization

For AI-based optimization, the surrogate serves as a data-generation vessel, meaning that optimization policies (e.g., SOA operating point, gain/saturation trade-offs, and receiver/ADC co-design choices) will only generalize if the impairment map is physically correct. Calibration hooks serve two critical roles:

- They enforce *physical plausibility constraints* across coupled effects (gain, saturation, and ASE), preventing ML from exploiting unrealistic combinations that would never occur in hardware.
- They enable *domain transfer* from lab measurements to simulation: by anchoring the surrogate parameters to measured gain curves and OSNR/BER trends, the training data inherits the same operating-point sensitivities that will appear in real experiments.

From an engineering perspective, these hooks enable the SOA block to be reused as if it were a real device across multiple setups (such as today’s single-pol baseline and future extended scenarios), while maintaining traceability between each model parameter and a measurable quantity. This is essential when the surrogate is used to inform design decisions rather than just to demonstrate qualitative understanding behavior³⁷.

6.2 SOA Operating Regimes and Design Trade-offs

This section maps the SOA model from Section 6.1 to *operating regimes* and *design degrees of freedom* that directly influence IM/DD performance. This defines the controllable parameters accessible to the digital surrogate and AI-driven optimization. In amplified IM/DD links, the SOA gain setting (along with the associated noise figure and filtering) collectively determines whether performance is limited by receiver noise, mixed noise and distortion, or ASE-related beat-noise processes. Therefore, the “best” operating point is not a fixed device parameter but

³⁷ M. S. Faruk and S. Savory, “Measurement Informed Models and Digital Twins for Optical Fiber Communication Systems”, Journal of Lightwave Technology, vol. 42(3), pp. 1016-1030, 2024.

a system-level optimum that must be identified based on the actual bandwidth and modulation constraints³⁸.

6.2.1 Gain-noise trade-off and optimum operating point

Optical pre-amplification enhances receiver sensitivity mainly by increasing the electrical signal level detected, making downstream thermal and shot noise less significant. This results in a larger usable power budget and better reach margins in IM/DD access-type links³⁹. However, the same gain that raises the signal also raises the amplified-spontaneous-emission (ASE) contribution generated by the amplifier, so the effective benefit of increasing gain eventually saturates and can reverse once ASE-related terms dominate. In system terms, SOA operation is therefore governed by a gain-noise trade-off: a low-gain regime where increasing gain is beneficial (receiver-noise-limited), an intermediate regime where both receiver noise and ASE-related beat noise contribute, and a high-gain regime where “extra gain is harmful” because ASE-driven noise (and its beating products after detection) sets the electrical SNR/BER⁴⁰.

A concise way to describe the design logic is to consider the optimal point where the incremental reduction in receiver-noise impact from increasing gain equals or is less than the incremental increase in ASE-driven noise at the decision variable. This is the practical meaning behind “*more gain is not always better*” in optically preamplified direct detection: once the signal-ASE and ASE-ASE beat-noise terms (as shaped by optical and electrical filtering) dominate the noise variance at the sampler, additional gain yields negligible SNR improvement while

³⁸ A. Omid, M. Banawan, E. Weckenmann, B. Paquin, A. Geravand, Z. Zheng, W. Shi, M. Zeng, and L. Rusch, “Auto-Encoder Optimized PAM IM/DD Transceivers for Amplified Fiber Links”, *Journal of Lightwave Technology*, vol. 43, pp. 547-556, 2024.

³⁹ S. Sarmiento, J. Mendinueta, J. Altabás, S. Spadaro, S. Shinada, H. Furukawa, J. Olmos, J. Lázaro, and N. Wada, “Optical Power Budget Enhancement in 50-90 Gb/s IM-DD PONs With NOMA-CAP and SOA-Based Amplification”, *IEEE Photonics Technology Letters*, vol. 32, pp. 608-611, 2020.

⁴⁰ J. Witzens, J. Müller, and A. Moscoso-Mártir, “Modification of Level Dependent ASE-Signal Beat Noise by Optical and Electrical Filtering in Optically Preamplified Direct Detection Receivers”, *IEEE Photonics Journal*, vol. 10(1), 7500416, 2018.

continuing to increase ASE contributions⁴¹. In our surrogate, this operating-point concept is important because the SOA gain is not merely a static link-budget knob: it is an *optimisable variable* whose optimal value depends on modulation statistics (e.g., multilevel PAM spacing) and receiver bandwidths, and it must be swept together with filtering choices to generate training data that preserves the correct sensitivity/OSNR trade-off landscape for AI⁴².

This perspective remains directly relevant to bipolar PAM (BPAM): because BPAM decisions depend on accurately maintaining relative level spacing around the reference (including the neighborhood of the “zero” crossing), excessive ASE-dominated operation compresses the effective electrical SNR and blurs the separation of adjacent decision regions, especially when the downstream receiver chain and equalization need to resolve smaller eye openings. In practice, the surrogate should therefore consider the SOA operating point as a constrained optimum, high enough to prevent receiver-noise dominance but limited so that ASE-related beat noise does not become the main impairment, so that AI-driven optimization learns a physically valid trade-off rather than a regime artificially “improved” by unrealistically increasing gain.

6.2.2 Input power dependence and saturation-induced distortion

In an SOA-preamplified IM/DD receiver, the instantaneous optical input power at the SOA, $P_{\text{in}}(t)$, is not merely a scaling factor; it directly determines whether the device operates in its linear (small-signal) regime or in saturation. As $P_{\text{in}}(t)$ increases, carrier depletion reduces the incremental gain, producing a compressive amplitude transfer that distorts multilevel modulation formats and progressively closes the eye. This is particularly relevant in our surrogate because the receiver-side SOA is treated as a *system degree of freedom* (operating point), and the input power is a *random process* shaped by modulation, bandwidth limitation, and noise;

⁴¹ *Ibidem*.

⁴² A. Omid, M. Banawan, E. Weckenmann, B. Paquin, A. Geravand, Z. Zheng, W. Shi, M. Zeng, and L. Rusch, “Auto-Encoder Optimized PAM IM/DD Transceivers for Amplified Fiber Links”, *Journal of Lightwave Technology*, vol. 43, pp. 547-556, 2024.

therefore, the link can move between linear and nonlinear regimes on a symbol-by-symbol basis when operating near saturation⁴³.

A standard system-level representation of saturation is to model the (power) gain as a decreasing function of the input power, e.g., in reduced form

$$G(P_{\text{in}}) = \frac{G_0}{1 + P_{\text{in}}/P_{\text{sat,in}}} \quad \Leftrightarrow \quad G_{\text{dB}}(P_{\text{in}}) = G_{0,\text{dB}} - 10 \log_{10} \left(1 + \frac{P_{\text{in}}}{P_{\text{sat,in}}} \right),$$

or via a gain – compression factor ε (often written against output power) that captures the same physical effect through a first-order carrier-depletion abstraction. In all these formulations, the “3-dB gain compression point” corresponds to the input (or output) power for which the gain is reduced by 3 dB from its small-signal value, and it provides an engineering threshold beyond which nonlinear distortion rapidly becomes dominant. These reduced models are widely used in communication-oriented SOA simulations because they preserve the correct qualitative behavior (linear and compressive) while remaining parameterizable from datasheets and measurements, and they map directly to system metrics such as eye opening and BER penalties⁴⁴.

From a waveform perspective, the saturation mechanism introduces amplitude-dependent gain:

$$P_{\text{out}}(t) = G(P_{\text{in}}(t))P_{\text{in}}(t),$$

so larger instantaneous samples experience smaller incremental gain. Even when the average input power is below the compression point, the peaks of $P_{\text{in}}(t)$, which become more pronounced as the symbol rate increases and pulse shaping creates overshoots. It can temporarily push the SOA into the nonlinear region. The resulting distortion is not “white” and not symmetric: it concentrates on high-amplitude

⁴³ V. Belykh, J. Rautert, V. Mikhlin, and A. Gubenko, “Bit Error Rate in WDM Data Transmission Links With Semiconductor Optical Amplifier”, *Journal of Lightwave Technology*, 43, 1076-1088, 2025.

⁴⁴ O.M. Kharraz, A. Supa’at, A. Atieh, A. Zamzuri, and M. Mahdi, “Enhanced SOA Gain Saturation Model for WDM Fiber Optic Transmission Systems”, *IET Optoelectronics*, vol 12(6), 263-268, 2018.

excursions, modifies the distribution of sample values, and changes the optimum decision thresholds at the receiver. In the digital surrogate, this is exactly why we expose input power and gain-compression parameters as controllable quantities: they determine whether the SOA acts as an approximately linear preamplifier (beneficial) or as a nonlinear limiter (harmful) for a given modulation and baud rate⁴⁵.

This effect becomes more severe for multilevel PAM because the constellation is encoded in *amplitude spacings* that must remain distinct after the optical front end. Consider an ideal 4-PAM intensity waveform with four target levels ($P_0 < P_1 < P_2 < P_3$). Under gain compression, each level is mapped through a different effective gain:

$$P'_k = G(P_k)P_k, \quad k \in 0,1,2,3,$$

So the *post-SOA* level spacings $\Delta'_k = P'_{k+1} - P'_k$ become non-uniform and typically shrink more at the upper levels (compression). In an eye diagram, this appears as a reduced vertical opening for the “upper eyes” and increased overlap between adjacent symbol rails; in a BER framework, it increases the likelihood that noise and residual ISI move samples across now-closer decision thresholds. This is precisely why “launch/input power optimization” is essential in SOA-assisted PAM links: raising P_{in} improves electrical SNR up to a point, but once compression is engaged, it reduces the *effective* minimum distance between levels and can worsen performance even if OSNR improves⁴⁶.

For BPAM-oriented operation (where the transmitter uses biasing and modulation to enable bipolar signaling concepts in an IM/DD setting), the same principle applies: any mechanism that nonlinearly reshapes amplitude statistics, especially compressive distortion around high peaks, directly affects the

⁴⁵ V. Belykh, J. Rautert, V. Mikhlin, and A. Gubenko, “*Bit Error Rate in WDM Data Transmission Links With Semiconductor Optical Amplifier*”, Journal of Lightwave Technology, 43, 1076-1088, 2025.

⁴⁶ M. Heck, “*Energy-efficient semiconductor optical amplifier with longitudinally varying confinement factor*”, Optics Letters, 50 (2), 301-304, 2025.

separability of decision regions and can reduce eye opening. In practice, BPAM waveforms may also exhibit different peak-to-average characteristics and occupy low-power regions differently from traditional unipolar PAM, which can alter how often the SOA is driven near compression and how strongly carrier-depletion dynamics imprint pattern-related distortion. Therefore, the surrogate’s ability to vary input power and SOA operating point while recording eye metrics provides a crucial basis for future BPAM-specific optimization studies, including robust biasing strategies and data-driven equalization under realistic front-end nonlinearities⁴⁷.

Finally, saturation-induced distortion is not only a limitation but also an *opportunity* for DSP/AI-assisted mitigation, provided the surrogate captures the nonlinearity with sufficient fidelity. Two representative mitigation directions are: (i) *digital compensation* that explicitly targets the nonlinear SOA transfer characteristic (predistortion or post-compensation) to restore quasi-linear level spacing, and (ii) *sequence-aware equalization* (e.g., recurrent or Volterra-style structures) that can learn the effective nonlinear channel when compression and bandwidth interact. In our framework, modeling saturation at the correct insertion point enables the generation of training data in which the impairment is *causal*, *amplitude-dependent*, and *operating-point-dependent*, which is essential for ML optimization to converge to physically meaningful solutions rather than “overfitting” an unrealistically linear front end⁴⁸.

6.2.3 OSNR and eye metrics under different gain settings

In optically pre-amplified IM/DD reception, OSNR acts as the key parameter that connects the optical amplifier settings (gain, operating point) to the electrical penalties observed at the sampler: increased amplitude variance, decreased eye opening, and ultimately BER degradation. In practice, OSNR is typically measured

⁴⁷ V. Belykh, J. Rautert, V. Mikhlin, and A. Gubenko, “*Bit Error Rate in WDM Data Transmission Links With Semiconductor Optical Amplifier*”, *Journal of Lightwave Technology*, 43, 1076-1088, 2025.

⁴⁸ K. Kaje, M. Al-Qadi, and R. Hui, “*Digital Compensation of SOA-Induced Nonlinearities in Field-Modulated Direct-Detection Systems*”, *IEEE Photonics Journal*, 13(6), 7200405, 2021.

with an OSA using a reference optical bandwidth (commonly $\Delta\lambda = 0.1\text{nm}$ around 1550 nm), so that the reported value reflects a ratio between the in-band signal power and the ASE power collected in that reference window. In SOA-based pre-amplified links, adjusting the SOA gain affects both the amplified signal power and the in-band ASE content⁴⁹; therefore, OSNR at the receiver is not a monotonic proxy for “better” performance, and must be considered together with saturation and pattern effects already discussed in Secs. 6.2.1-6.2.2.

A convenient starting point is the operational OSNR definition

$$OSNR_{\Delta\lambda} = \frac{P_{\text{sig}}}{P_{\text{ASE}}(\Delta\lambda)},$$

$$OSNR_{\Delta\lambda}(\text{dB}) = 10 \log 10 \left(\frac{P_{\text{sig}}}{P_{\text{ASE}}(\Delta\lambda)} \right),$$

where $P_{\text{ASE}}(\Delta\lambda)$ is the ASE power integrated over the OSA reference bandwidth. For implementation and calibration, it is often useful to convert the OSA reference window into an equivalent optical noise bandwidth in Hz, using the first-order mapping

$$\Delta f \approx \frac{c}{\lambda^2} \Delta\lambda,$$

so that an ASE spectral density estimate can be formed as $S_{\text{ASE}} \approx P_{\text{ASE}}(\Delta\lambda)/\Delta f$. This is particularly relevant in a digital surrogate because the subsequent receiver chain effectively applies its own optical/electrical filtering, meaning that the “effective” noise that drives eye closure is the ASE integrated over the surrogate’s modeled optical filter and receiver front-end bandwidth rather than over the OSA reference window. Consistency between $\Delta\lambda$ -based OSNR reporting and the surrogate’s internal bandwidth bookkeeping is essential if the surrogate is to

⁴⁹ V. Belykh, J. Rautert, V. Mikhlin, and A. Gubenko, “Bit Error Rate in WDM Data Transmission Links With Semiconductor Optical Amplifier”, *Journal of Lightwave Technology*, vol. 43(3), pp. 1076-1088, 2025.

reproduce lab-measured OSNR sweeps and generate ML training data that remains valid under different bandwidth settings⁵⁰.

From a system-level viewpoint, the SOA gain setting controls OSNR through two coupled mechanisms: (i) amplification of the signal component and (ii) addition of amplifier noise that occupies the signal band. In other words, increasing G increases $P_{\text{sig,out}}$ but also increases $P_{\text{ASE,out}}$, and the latter can dominate beyond an optimum operating region. This explains why, in pre-amplified IM/DD experiments and simulations, one typically observes a *BER-versus-OSNR “knee”* and a corresponding *eye-height/Q-factor saturation* when gain is increased beyond the point at which ASE and/or saturation-induced distortion becomes limiting. Such behaviors have been reported in recent studies of SOA-based IM/DD systems that analyze performance as a function of OSNR and amplifier operating point⁵¹.

To translate OSNR into waveform-quality indicators for multilevel formats (PAM4 and BPAM), the surrogate uses *eye-based metrics* computed at the decision sampling instant. For an (M)-level signal, denote by μ_k and σ_k the conditional mean and standard deviation of the received samples associated with the k -th level (after timing recovery/sampling alignment). A practical definition of *local* eye quality between adjacent levels k and $k + 1$ is

$$Q_k = \frac{\mu_{k+1} - \mu_k}{\sigma_{k+1} + \sigma_k}, \quad Q_{\min} = \min_k Q_k$$

and a corresponding minimum eye height proxy is

$$H_{\text{eye,min}} = \min_k (\mu_{k+1} - \mu_k),$$

⁵⁰ D. Lu, X. Zhou, Y. Yang, J. Huo, J. Yuan, K. Long, C. Yu, A. Lau, and C. Lu, “*Theoretical analysis of PAM-N and M-QAM BER computation with single-sideband signal*”, Science China Information Sciences, vol. 64, 182312, 2021.

⁵¹ V. Belykh, J. Rautert, V. Mikhlin, and A. Gubenko, “*Bit Error Rate in WDM Data Transmission Links With Semiconductor Optical Amplifier*”, Journal of Lightwave Technology, vol. 43(3), pp. 1076-1088, 2025.

with penalties interpreted based on how OSNR-driven noise increases and reduces the remaining usable separations for decision thresholds. This formulation is directly applicable to PAM4 and also naturally extends to BPAM once the decision regions are expressed in terms of the effective level mapping at the photodetector output (i.e., the OSNR-driven fluctuations decrease the separability of the “polarity-resolved” levels). This eye/Q-factor framing is widely used in system papers that evaluate SOA-assisted direct-detection links under OSNR variation and report eye/Q/BER trends⁵².

6.2.4 EDFA vs SOA as pre-amplifiers in high-speed IM/DD

Optical pre-amplification is commonly used in high-speed IM/DD experiments because it reduces the receiver sensitivity requirement and allows for controlled OSNR sweeps. However, the type of pre-amplifier significantly influences the impairment profile experienced by the detector and the DSP. EDFAs are frequently employed in laboratory demonstrations because they are mature, readily available, and can deliver high, spectrally flat gain with good OSNR characteristics over the C-band. This simplifies experimental link-budget closure and benchmarking⁵³. In access-oriented IM/DD studies, EDFA boosting is often used as a convenient baseline for “amplified reception” because it enables the system to be evaluated under well-controlled gain and noise conditions with minimal additional nonlinear penalties from the amplifier itself⁵⁴.

SOAs, in contrast, are compelling when the target system emphasizes *integration and deployment realism* rather than laboratory convenience. Their key advantage is that they can be co-packaged or monolithically integrated with

⁵² K. Kaje, M. Al-Qadi, and R. Hui, “*Digital Compensation of SOA-Induced Nonlinearities in Field-Modulated Direct-Detection Systems*”, IEEE Photonics Journal, vol. 13(6), 7200405, 2021.

⁵³ S. Hayle, H. Hsu, C. Wang, H. Lu, J. Lu, W. Hsu, Y. Chung, Y. Bai, and K. Okram, “*High-speed FSO-5G wireless communication system with enhanced loss compensation using high-power EDFA*”, Scientific Reports, vol. 15, 2025.

⁵⁴ S. Sarmiento, J. Mendinueta, J. Altabás, S. Spadaro, S. Shinada, H. Furukawa, J. Olmos, J. Lázaro, and N. Wada, “*Optical Power Budget Enhancement in 50-90 Gb/s IM-DD PONs With NOMA-CAP and SOA-Based Amplification*”, IEEE Photonics Technology Letters, vol. 32, pp. 608-611, 2020.

photonic receivers, enabling compact and potentially cost-effective amplified front-ends for next-generation PONs and short-reach links where footprint, power, and integration are critical constraints⁵⁵. Recent high-dynamic-range PON demonstrations explicitly exploit SOA pre-amplification to push power budgets while maintaining high symbol rates, illustrating why SOAs are increasingly relevant for IM/DD receiver architectures despite their more complex impairment landscape⁵⁶.

From a noise perspective, both EDFAs and SOAs add ASE, but their system-level consequences differ because the receiver ultimately performs square-law detection, converting optical noise into electrical noise through beating terms. Classical optically pre-amplified direct-detection theory shows that the dominant noise contributions at the photodetector can include signal-ASE and ASE-ASE beat noise, and the relative strength of these terms depends on the amplifier noise properties and on the optical/electrical filtering used in the receiver⁵⁷. In practice, SOA-based pre-amplified receivers are often more sensitive to careful bandwidth engineering because the *level-dependent* redistribution of signal-ASE beat noise by optical and electrical filtering can directly reshape PAM eye openings and Q metric, i.e., the filter chain is not a passive “cleanup” stage but an active determinant of the effective decision-variable statistics⁵⁸. For this reason, comparisons between EDFA and SOA pre-amplification must be made in a measurement-consistent way: OSNR defined in an OSA reference bandwidth is an important reporting metric, but the

⁵⁵ A. Reza, L. Venkatasubramani, and L. Barry, “112.5 Gbit/s long reach passive optical network with over 31 dB power budget enabled by semiconductor optical amplifiers”, *Scientific Reports*, vol. 15, 2025.

⁵⁶ S. Murphy, F. Jamali, P. Townsend, and C. Antony, “High Dynamic Range 100G PON Enabled by SOA Preamplifier and Recurrent Neural Networks”, *Journal of Lightwave Technology*, vol. 41(11), pp. 3522-3532, 2023.

⁵⁷ G. Cariolaro, R. Corvaja, and G. Pierobon, “Exact performance evaluation of lightwave systems with optical preamplifier”, *European Transactions on Telecommunications*, vol. 5, pp. 757-766, 1994.

⁵⁸ J. Witzens, J. Müller, and A. Moscoso-Mártir, “Modification of Level Dependent ASE-Signal Beat Noise by Optical and Electrical Filtering in Optically Preamplified Direct Detection Receivers”, *IEEE Photonics Journal*, vol. 10, 2018.

electrical-domain penalties are ultimately set by how much ASE falls within the receiver's effective noise-equivalent bandwidth and how it beats with the signal after detection⁵⁹.

Beyond noise, SOAs introduce a second class of impairments that is comparatively weak (or absent) for EDFAs in the IM/DD regimes considered here: gain compression and carrier-dynamics effects. In high-speed multilevel PAM, SOA saturation compresses upper amplitude excursions more strongly than lower ones, reducing effective level spacing and creating deterministic distortion that manifests as eye closure and threshold sensitivity (as developed in Sec. 6.2.2)⁶⁰. Moreover, because SOA gain recovery occurs on finite carrier timescales, the instantaneous gain can become pattern dependent, producing a data-dependent channel with memory (data-dependent ISI) that can be especially relevant in burst-mode PON operation and at very high symbol rates⁶¹. These SOA-specific nonlinear and memory effects motivate the use of nonlinear equalization and sequence-aware learning approaches; for example, recurrent neural network equalizers have been demonstrated as effective tools to mitigate SOA-induced nonlinearities in high-baud-rate PAM IM/DD WDM-PON settings, directly linking SOA operating regimes to DSP/ML strategy selection⁶².

EDFAs, by comparison, are often treated as effectively linear (with respect to the signal's intensity swings) under typical IM/DD conditions, so the amplified link behaves more like a "noise-shaped but memory-light" receiver problem: OSNR and filtering dominate, while amplifier-induced pattern effects are not a primary limitation in the same sense as with SOAs. This distinction is crucial when

⁵⁹ *Ibidem*.

⁶⁰ V. Belykh, J. Rautert, V. Mikhlin, and A. Gubenko, "Bit Error Rate in WDM Data Transmission Links With Semiconductor Optical Amplifier", *Journal of Lightwave Technology*, vol. 43, pp. 1076-1088, 2025.

⁶¹ S. Murphy, F. Jamali, P. Townsend, and C. Antony, "High Dynamic Range 100G PON Enabled by SOA Preamplifier and Recurrent Neural Networks", *Journal of Lightwave Technology*, vol. 41(11), pp. 3522-3532, 2023.

⁶² A. Reza, M. Troncoso-Costas, C. Browning, S. O'duill, and L. Barry, "Mitigation of SOA-Induced Nonlinearities With Recurrent Neural Networks in 75 Gbit/s/λ PAM-4 IM/DD WDM-PON Transmission Systems", *Journal of Lightwave Technology*, vol. 41, pp. 3967-3975, 2023.

interpreting EDFA-based experimental baselines against SOA-oriented deployment scenarios: equal OSNR does not guarantee equal decision statistics if one pre-amplifier introduces operating-point-dependent compression and memory while the other does not⁶³.

Finally, there is a practical device-level consideration that becomes system-relevant in modular receiver design: polarization dependence. EDFAs are typically treated as polarization-insensitive for system modeling purposes in IM/DD links, whereas SOAs may exhibit polarization-dependent gain, meaning that the effective amplification can vary with the state of polarization at the SOA input. In tightly controlled single-polarization lab operation this can be neglected, but in broader reuse of the same SOA “device-equivalent” model, e.g., for different receiver modules or less-controlled fiber conditions, polarization dependence becomes a potential penalty mechanism that must be parameterized (or bounded) to preserve model transferability⁶⁴.

In the context of our digital surrogate, these EDFA-SOA distinctions are not merely descriptive; they define the optimization landscape that ML methods will see. EDFA-like pre-amplification corresponds primarily to a noise-limited regime with filtering-consistent OSNR mapping, while SOA pre-amplification adds operating-point-dependent nonlinear distortion and memory that can reshape the optimal bias/gain/power settings and the choice of equalization strategy. The surrogate’s strength is precisely its ability to represent these phenomena in a controlled and modular way: by sweeping gain and input power through linear and saturated regimes and by generating ASE-consistent noise statistics (under receiver bandwidth constraints), the model can produce training data that exposes ML algorithms to the same trade-offs that dominate real IM/DD receiver design. This is aligned with recent work showing that end-to-end learning or receiver-side ML

⁶³ G. Cariolaro, R. Corvaja, and G. Pierobon, “*Exact performance evaluation of lightwave systems with optical preamplifier*”, European Transactions on Telecommunications, vol. 5, pp. 757-766, 1994.

⁶⁴ A. Reza, L. Venkatasubramani, and L. Barry, “*112.5 Gbit/s long reach passive optical network with over 31 dB power budget enabled by semiconductor optical amplifiers*”, Scientific Reports, vol. 15, 2025.

can exploit accurate amplifier-aware models to optimize IM/DD performance under amplified-link constraints⁶⁵.

6.3.1 Photodetection, responsivity, and PD vs TIA conversion

In an IM/DD receiver, the optical waveform reaching the receiver front-end is represented by the instantaneous optical power $P(t)$ at the photodetector input. Under the standard small-signal/linear operating assumption of a PIN photodiode (i.e., no space-charge-limited regime and no saturation in the electrical chain), the detected signal current is modeled as proportional to optical power through the responsivity $\mathcal{R}$ ([A/W]):

$$i_{\text{sig}}(t) = \mathcal{R}P(t).$$

This relation is the canonical starting point for system-level IM/DD modeling. It is also the basis of high-speed PAM4 experimental and modeling studies, where $\mathcal{R}$ is treated as a measurable device parameter specified at the operating wavelength⁶⁶.

In the digital surrogate, the photodetection block is designed to be interface-consistent with upstream optical modules: the photoreceiver input is naturally expressed as a power waveform $P(t)$ in watts, which directly maps to current by $i(t) = \mathcal{R}P(t)$. For completeness, an optional interface can accept a complex optical field representation $E(t)$ and internally construct power through the square-law mapping $P(t) = |E(t)|^2$ before applying responsivity. The underlying scientific meaning remains unchanged: the electrical decision variable in IM/DD is generated from optical intensity, and the surrogate therefore treats photodetection as an

⁶⁵ A. Omid, M. Banawan, E. Weckenmann, B. Paquin, A. Geravand, Z. Zheng, W. Shi, M. Zeng, and L. Rusch, “Auto-Encoder Optimized PAM IM/DD Transceivers for Amplified Fiber Links”, *Journal of Lightwave Technology*, vol. 43, pp. 547-556, 2024.

⁶⁶ Z. Feng, N. Li, W. Li, P. He, M. Luo, Q. Hu, L. Huang, and Y. Jiang, “A Simplified Volterra Equalizer Based on System Characteristics for Direct Modulation Laser (DML)-Based Intensity Modulation and Direct Detection (IM/DD) Transmission Systems”, *Photonics*, 10(10), 1174, 2023.

intensity-to-current transduction stage whose validity is governed by linearity assumptions and parameter calibration (e.g., $\mathcal{R}$ at the operating wavelength)⁶⁷.

After the photocurrent is generated, the surrogate supports two distinct but physically meaningful electrical conversion configurations, reflecting common receiver implementations:

1. **Photodiode + load (PD mode).** The photocurrent is terminated on an effective load R_{load} and converted to a voltage,

$$v_{\text{out}}(t) = R_{\text{load}}i(t),$$

which is a standard abstraction for direct electrical termination or simplified receiver front-ends.

2. **Photoreceiver / TIA chain (PR mode).** The photocurrent is amplified by a transimpedance amplifier (TIA). At the system level, the TIA is modeled as a linear transimpedance response $Z_{\text{TIA}}(f)$ (or impulse response $z_{\text{TIA}}(t)$), such that

$$V_{\text{out}}(f) = Z_{\text{TIA}}(f)I(f), \quad v_{\text{out}}(t) = (z_{\text{TIA}} * i)(t).$$

This representation captures the fundamental role of the TIA: converting current into a voltage waveform suitable for multilevel detection, while introducing the practical design trade-offs between transimpedance gain, bandwidth, and noise⁶⁸.

The choice between these configurations is not merely a matter of implementation convenience. It controls the *electrical scaling and the effective receiver front-end abstraction* used when generating ML-grade datasets. High-speed IM/DD receivers frequently rely on close co-design and/or integration of the PD and the TIA to manage parasitics and achieve the gain-bandwidth-noise targets

⁶⁷ A. Yuan and S. Savory, “Data-Driven Modeling of a Direct Detection System With an Electro-Absorption Modulated Laser”, *Journal of Lightwave Technology*, vol. 43, pp. 5545-5557, 2025.

⁶⁸ K. Li, S. Liu, X. Ruan, D. Thomson, Y. Hong, F. Yang, L. Zhang, C. Lacava, F. Meng, W. Zhang, P. Petropoulos, F. Zhang, and G. Reed, “Co-design of a differential transimpedance amplifier and balanced photodetector for a sub-pJ/bit silicon photonics receiver”, *Optics Express*, vol. 28, no. 9, pp. 14038-14054, 2020.

required by multilevel formats. This is particularly relevant in PAM4 links where decision thresholds depend on accurate preservation of level spacing in the electrical domain. In this context, integrated or co-designed PD/TIA receivers motivate modeling the current-to-voltage conversion as a dedicated transimpedance stage rather than a pure resistive load.

Finally, it is important to emphasize what this subsection establishes for the remainder of Sec. 6.3: the surrogate's PD/PR block produces an electrical waveform whose amplitude statistics (mean rail locations and spreads) are subsequently shaped by (i) signal-dependent and signal-independent noise terms (Sec. 6.3.2) and (ii) bandwidth limitation/receiver filtering (Sec. 6.3.3). In other words, Sec. 6.3.1 defines the deterministic electrical mapping ($P(t) \rightarrow i(t) \rightarrow v(t)$), which is the reference on which the physically correct noise and bandwidth constraints will act⁶⁹.

6.3.2 Shot noise, dark current, and thermal/TIA noise terms (PSD and variance)

In an IM/DD receiver, the deterministic photocurrent $i_{\text{sig}}(t) = \mathcal{R}P(t)$ is unavoidably accompanied by stochastic fluctuations originating from the discreteness of charge transport (shot noise) and from electrical noise generated in the receiver front-end (thermal and amplifier/TIA noise). From a system viewpoint, these noise terms directly shape the level-dependent decision statistics of multilevel PAM (e.g., PAM4/BPAM), because the conditional rail distributions at the sampler output are characterized by a mean μ_k and a variance σ_k^2 that are influenced by both signal-dependent and signal-independent noise sources. Consequently, a physically consistent noise budget is essential in the digital surrogate to produce datasets that

⁶⁹ Z. Feng, N. Li, W. Li, P. He, M. Luo, Q. Hu, L. Huang, and Y. Jiang, “A Simplified Volterra Equalizer Based on System Characteristics for Direct Modulation Laser (DML)-Based Intensity Modulation and Direct Detection (IM/DD) Transmission Systems”, *Photonics*, 10(10), 1174, 2023.

remain faithful to measured BER/eye behavior and to enable reliable AI-based optimization of operating points and DSP strategies⁷⁰.

Shot noise (signal-dependent) and the role of dark current.

Shot noise is commonly modeled through a single-sided current-noise power spectral density (PSD) proportional to the *total* average photodiode current,

$$S_{i,\text{shot}} = 2qI_{\text{tot}} \quad [\text{A}^2/\text{Hz}],$$

where q is the elementary charge and I_{tot} is the sum of the signal-generated current and any background current contributions (including dark current). For a receiver chain whose effective noise bandwidth is B_n , the corresponding mean-square shot-noise current is

$$\langle i_{\text{shot}}^2 \rangle = 2qI_{\text{tot}}B_n, \quad \sigma_{\text{shot}} = \sqrt{2qI_{\text{tot}}B_n}.$$

This formulation is particularly important for IM/DD multilevel formats because I_{tot} changes with the transmitted symbol level, making the shot-noise variance inherently level-dependent⁷¹.

Thermal noise (signal-independent) referenced to resistive elements.

Thermal noise originates from resistive elements in the receiver front-end (e.g., load resistance, feedback resistance, and any explicit resistive terminations). In the current form, the single-sided thermal-noise PSD of a resistor R at temperature T is

⁷⁰ G. Rizzelli, P. Torres-Ferrera, F. Forghieri, and R. Gaudino, “An Analytical Model for Performance Estimation in Modern High-Capacity IMDD Systems”, *Journal of Lightwave Technology*, vol. 42, pp. 1443-1452, 2024.

⁷¹ Y. Jauregui-Sánchez, P. Clemente, P. Latorre-Carmona, E. Tajahuerce, and J. Lancis, “Signal-to-noise ratio of single-pixel cameras based on photodiodes”, *Applied Optics*, vol. 57, no. 7, pp. B67-B73, 2018.

$$S_{i,\text{th}} = \frac{4kT}{R} \quad [\text{A}^2/\text{Hz}],$$

and integrating over the receiver noise-equivalent bandwidth B_n yields

$$\langle i_{\text{th}}^2 \rangle = \frac{4kT}{R} B_n, \quad \sigma_{\text{th}} = \sqrt{\frac{4kT}{R} B_n}.$$

Equivalently, the same physics can be expressed in voltage form as $\langle v_{\text{th}}^2 \rangle = 4kTRB_n$, depending on whether the receiver is modeled as current- or voltage-referred. This conversion flexibility is useful in a surrogate because a PD-only termination and a TIA-based photoreceiver naturally suggest different “reference planes” for the noise budget⁷².

TIA noise densities and input-or output-referred interpretation.

In practical photoreceivers, the transimpedance amplifier introduces additional internal noise contributions that are often summarized using (i) an *input-referred current-noise density* $i_n [\text{A}/\sqrt{\text{Hz}}]$ and (ii) an *output-referred voltage-noise density* $e_n [\text{V}/\sqrt{\text{Hz}}]$. At a first-order system level, the resulting output RMS noise (before any additional filtering) can be treated as the bandwidth-integrated combination of an input-current noise amplified by the transimpedance and a directly-added output-voltage noise, i.e.,

$$\sigma_{v,\text{out}}^2 \approx |Z_{\text{TIA}}|^2 i_n^2 B_n + e_n^2 B_n,$$

with refinements introduced when the full frequency dependence $Z_{\text{TIA}}(f)$ and noise-gain shaping are explicitly modeled. This input- or output-referred viewpoint is particularly compatible with a modular digital surrogate: it allows the same photoreceiver abstraction to be calibrated either from component-level

⁷² K. Achtenberg and Z. Bielecki, “Comparative Analysis of the Selected Photoreceiver Input Stages in Terms of Noise”, *Sensors*, 25(5), 1359 2025.

specifications (noise densities, transimpedance gain, bandwidth) or from measured electrical noise floors⁷³.

Noise-equivalent bandwidth and discrete-time implementation consistency.

A key modeling nuance is that all PSD-based expressions above become variances only after consistent bandwidth integration. The *noise-equivalent bandwidth* B_n is defined as the equivalent rectangular bandwidth that produces the same integrated noise power as the actual receive filter. For a first-order low-pass response, B_n is commonly related to its 3-dB bandwidth $f_{3\text{dB}}$ by

$$B_n \approx \frac{\pi}{2} f_{3\text{dB}},$$

which provides a convenient bridge between an electrical one-pole bandwidth abstraction and the expressions for thermal and shot noise variance. In the surrogate, using a consistent B_n is essential because it ensures that (i) shot noise scales correctly with received power and symbol level, and (ii) thermal/TIA noise scales correctly with receiver bandwidth, enabling meaningful comparisons across different sampling rates, front-end bandwidth settings, and equalization strategies⁷⁴.

Mapping into the surrogate (PD vs photoreceiver mode) and relevance to ML optimization.

The implemented receiver block follows the standard engineering structure: noise sources are formed in the natural domain where they originate (photocurrent-domain for shot noise and resistor-referred noise, plus additional TIA noise-density terms in photoreceiver mode), integrated over a consistent effective bandwidth, and then mapped into the output voltage through the chosen current-to-voltage conversion path. This separation is not merely cosmetic: for bandwidth-limited IM/DD systems, the interaction among front-end noise, equalization, and “noise enhancement” can materially affect the post-DSP SNR and the rail-dependent

⁷³ Y. Yang, D. Li, N. Qi, and B. Wang, “A 2 pA/√Hz Input-Referred Noise TIA in 180-nm CMOS With 2.5 GHz Bandwidth for Optical Receiver”, IEEE Solid-State Circuits Letters, vol. 8, pp. 189-192, 2025.

⁷⁴ K. Achtenberg and Z. Bielecki, “Comparative Analysis of the Selected Photoreceiver Input Stages in Terms of Noise”, Sensors, 25(5), 1359 2025.

spreads, which directly impact BER predictions and the fidelity of ML-generated training data⁷⁵.

6.3.3 Receiver bandwidth limitation, one-pole abstraction, and ISI/eye closure

In high-speed IM/DD links (PAM4/BPAM), the photodiode-TIA front-end often acts as the primary bandwidth bottleneck. From a system perspective, the combined effect of the photodiode junction capacitance, input impedance, and TIA feedback network can be represented as a receive-side low-pass transfer function $H_{\text{RX}}(f)$ that shapes both the signal and the noise. This is an important detail: when the effective electrical bandwidth becomes comparable to (or less than) the symbol-rate spectral occupancy, the receiver response creates deterministic memory (ISI) and colored-noise effects that directly impact eye opening and achievable BER⁷⁶.

One-pole low-pass model and its time-domain implication (ISI “memory”)

A widely used first-order abstraction for bandwidth limitation is the one-pole low-pass model

$$H_{\text{RX}}(f) = \frac{1}{1 + j\frac{f}{f_c}}, \quad f_c = f_{3\text{dB}},$$

with the corresponding impulse response

$$h_{\text{RX}}(t) = \frac{1}{\tau} e^{-t/\tau} u(t), \quad \tau = \frac{1}{2\pi f_{3\text{dB}}}.$$

The exponential tail of $h_{\text{RX}}(t)$ implies that each symbol contributes residual energy to subsequent symbol intervals; therefore, the sampled waveform at the decision instants depends on a finite history of past symbols. In multilevel PAM,

⁷⁵ W. Wang, F. Li, D. Zou, W. Ni, Q. Sui, X. Yi, Y. Cai, and Z. Li, “*Suppressed Noise Enhancement Equalization Combined With Nonlinear-MLSE Techniques for Bandwidth-Limitation IM/DD System*”, Journal of Lightwave Technology, vol. 41, pp. 4655-4662, 2023.

⁷⁶ D. Li, L. Geng, F. Maloberti, and F. Svelto, “*Overcoming the Transimpedance Limit: A Tutorial on Design of Low-Noise TIA*”, IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 69, pp. 2648-2653, 2022.

this memory is particularly harmful because it produces level-dependent distortion: the conditional means μ_k of the four rails shift as a function of neighboring symbols, and the effective rail spacing shrinks non-uniformly, reducing the vertical eye openings and increasing decision errors even if the channel is otherwise linear⁷⁷.

From a receiver-design viewpoint, the same physical mechanism is often discussed as the *transimpedance gain-bandwidth limitation*: increasing transimpedance gain generally tightens the dominant pole set by the feedback network and input capacitances, so practical designs must explicitly trade gain, bandwidth, and noise. Modern high-speed PAM receivers therefore incorporate *co-optimization* of PD parasitics and TIA architecture to manage this gain-bandwidth constraint at the front-end, before any DSP can act⁷⁸.

Noise-equivalent bandwidth and “noise enhancement” under equalization

Bandwidth limitation impacts not only the deterministic response but also the *integrated noise power* at the sampler. In PSD-based modeling, the variance of an additive (approximately white) electrical noise term after filtering is proportional to the *noise-equivalent bandwidth*

$$B_n \triangleq \frac{1}{|H_{RX}(0)|^2} \int_0^\infty |H_{RX}(f)|^2 df,$$

so replacing an actual front – end by a one – pole $H_{RX}(f)$ provides a direct bridge between a hardware bandwidth parameter f_{3dB} and the variance used in receiver-level SNR/BER calculations. The key nuance for system simulations is that practical “bandwidth extension” techniques (e.g., RC compensation, peaking, CTLE-like equalization) reshape $H_{RX}(f)$ into a higher-order/non-flat magnitude

⁷⁷ M. Xiang, S. Fu, O. Xu, J. Li, D. Peng, Z. Gao, Y. Wang, and Y. Qin, “Advanced DSP Enabled C-Band 112 Gbit/s/λ PAM-4 Transmissions With Severe Bandwidth-Constraint”, Journal of Lightwave Technology, vol. 40, pp. 987-996, 2022.

⁷⁸ D. Patel, A. Sharif-Bakhtiar, and T. Carusone, “A 112-Gb/s —8.2-dBm Sensitivity 4-PAM Linear TIA in 16-nm CMOS With Co-Packaged Photodiodes”, IEEE Journal of Solid-State Circuits, vol. 58, pp. 771-784, 2023.

response: the apparent 3-dB bandwidth can increase. In contrast, the integrated noise (and its color) can also increase. Therefore, in the surrogate, the correct abstraction is not “ $f_{3\text{dB}}$ alone,” but the full $H_{\text{RX}}(f)$ (or an equivalent parametric model) that simultaneously determines ISI memory and noise integration⁷⁹.

This observation is key to data fidelity for ML optimization. If a surrogate uses an unrealistic flat or bandwidth-only model, it will misrepresent the trade-off between eye opening and noise increase, leading to datasets that can bias an ML optimizer toward operating points that cannot be reproduced in hardware. On the other hand, a physically consistent $H_{\text{RX}}(f)$ model allows the surrogate to produce training data where equalizer choices (FFE/DFE/MLSE, noise-whitening, etc.) are guided by actual channel memory and colored-noise statistics rather than idealized assumptions⁸⁰.

Eye closure under severe bandwidth constraints and the need for memory-aware DSP

When the end-to-end electrical bandwidth is severely limited (e.g., a few tens of GHz for symbol rates well above 50-100 GBaud), the received waveform becomes dominated by ISI: the eye diagram can appear nearly closed in the time domain, yet significant performance can still be recovered by explicitly modeling the front-end response as a finite-memory channel. In this regime, advanced DSP techniques (e.g., MLSE/BCJR-like sequence detection, channel shortening, or specialized noise-whitening methods) are designed to exploit the structured ISI induced by front-end bandwidth constraints and to reduce the impact of colored noise. This approach directly supports the surrogate’s goal by providing a controlled environment in which bandwidth, memory length, and equalization complexity can be systematically explored for PAM4/BPAM under IM/DD constraints⁸¹.

⁷⁹ M. Babar and G. Roberts, “A Multi-Stage RC Compensation Technique for Decoupling the Transimpedance and BW: Creating High Speed and Low Noise TIA Designs”, IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 72, pp. 3847-3860, 2025.

⁸⁰ W. Wang, F. Li, D. Zou, W. Ni, Q. Sui, X. Yi, Y. Cai, and Z. Li, “Suppressed Noise Enhancement Equalization Combined With Nonlinear-MLSE Techniques for Bandwidth-Limitation IM/DD System”, Journal of Lightwave Technology, vol. 41, pp. 4655-4662, 2023.

⁸¹ Q. Wu, Z. Xu, Y. Zhu, H. Ji, Y. Yang, G. Qiao, L. Liu, S. Wang, J. Liang, J. Wei, Z. He, Q. Zhuge, and W. Hu, “Gradient-descent noise whitening techniques for short-reach IM-DD optical

6.3.4 Complete PD+TIA front-end model tied to BER/Q/eye metrics

A physically consistent photodiode plus electrical front-end model is essential for an IM/DD digital surrogate because this component converts optical power into the electrical waveform, which ultimately influences eye opening, Q-factor, and BER. In modern high-speed PAM4 receivers, performance is often limited not by “ideal” detection but by a combination of front-end effects: (i) responsivity and conversion gain (which affect level spacing), (ii) both signal-dependent and signal-independent noise (leading to vertical eye closure), and (iii) limited electrical bandwidth (causing ISI/memory effects and colored noise). State-of-the-art silicon photonic receiver demonstrations highlight this interaction by reporting OMA sensitivity and achievable BER at very high lane rates, underscoring that the PD-TIA front-end is the practical bottleneck that must be modeled precisely for accurate predictions and ML dataset generation⁸².

(i) Unified block-level model: optical waveform to electrical waveform

In the surrogate, the PD front-end is abstracted as a deterministic conversion followed by additive noise and (optional) bandwidth shaping. The model accepts either optical power $P_{\text{in}}(t)$ or (when the preceding block outputs a complex field) an optical field $E(t)$ that is internally mapped to power via $P_{\text{in}}(t) = |E(t)|^2$ (unit-impedance convention consistent with the surrogate’s internal signal contracts). The photodiode conversion is then represented by the standard responsivity law

$$i_{\text{sig}}(t) = \mathcal{R}P_{\text{in}}(t),$$

which sets the mean level spacing at the electrical output after current-to-voltage conversion. This is the correct first-order starting point for IM/DD receiver

interconnects with severe bandwidth limitation”, Optics Express, vol. 32, no. 2, pp. 1715-1727, 2023.

⁸² D. Patel, A. Sharif-Bakhtiar, and T. Carusone, “A 112-Gb/s —8.2-dBm Sensitivity 4-PAM Linear TIA in 16-nm CMOS With Co-Packaged Photodiodes.”, IEEE Journal of Solid-State Circuits, vol. 58, pp. 771-784, 2023.

modeling and remains the implicit backbone even in advanced co-designed silicon photonics receivers that simultaneously push sensitivity and bandwidth⁸³.

To support two realistic experimental configurations, the surrogate implements two operating modes:

- **Photodiode-only (PD mode):** current is converted to voltage on a load resistor R_{load} ,

$$v_{\text{sig}}(t) = R_{\text{load}} i_{\text{sig}}(t).$$

- **Photoreceiver (PR mode):** current is converted by a TIA with transimpedance gain G_{TIA} (V/A),

$$v_{\text{sig}}(t) = G_{\text{TIA}} i_{\text{sig}}(t).$$

This distinction matters for a digital surrogate because it cleanly separates (a) a minimal physics-only PD termination from (b) a receiver-grade front-end whose noise densities, gain, and bandwidth are controllable degrees of freedom, precisely what is needed for parametric sweeps and ML-oriented data production.

(ii) Noise + bandwidth as the mechanism linking the front-end to eye/Q/BER

From the standpoint of eye and BER formation, the front-end produces a received voltage waveform

$$v_{\text{out}}(t) = \left(v_{\text{sig}}(t) + v_{\text{noise}}(t) \right) \xrightarrow{\text{front-end } H(f)} y(t),$$

where $H(f)$ represents the effective PD/TIA electrical transfer function (and any deliberate bandwidth shaping). High-speed receiver literature repeatedly shows that bandwidth is not merely a “loss”: it trades against integrated noise. It can be

⁸³ K. Li, S. Liu, X. Ruan, D. Thomson, Y. Hong, F. Yang, L. Zhang, C. Lacava, F. Meng, W. Zhang, P. Petropoulos, F. Zhang, and G. Reed, “Co-design of a differential transimpedance amplifier and balanced photodetector for a sub-pJ/bit silicon photonics receiver.”, Optics Express, vol. 28, no. 9, pp. 14038-14054, 2020.

intentionally designed as *sub-Nyquist* and then paired with equalization to optimize sensitivity at a target BER⁸⁴.

In our surrogate implementation, the noise is constructed to preserve the correct dependencies on the sampling rate and effective integration bandwidth. Specifically, the model uses an effective *noise-equivalent bandwidth* B_{eff} expressed as a fraction of the discrete-time Nyquist bandwidth,

$$B_{\text{eff}} = \text{enbwFrac} \cdot \frac{F_s}{2},$$

so that white-noise PSD terms become per-sample variances via multiplication by B_{eff} . This is crucial for ML dataset consistency: if F_s changes (e.g., due to different internal oversampling ratios), the injected noise statistics remain physically interpretable and scale correctly with bandwidth.

The surrogate then composes (depending on mode) the dominant noise contributions that determine the vertical eye opening:

- **Shot noise (signal + dark current):** constructed from the total diode current ($I_{\text{total}}(t) = i_{\text{sig}}(t) + I_{\text{dark}}$) and injected in the current domain before voltage conversion. This ensures the correct *signal dependence* of the noise amplitude, which is essential for multilevel PAM statistics.
- **Thermal noise (load/TIA-related):** modeled as an input-referred current noise term associated with resistive elements (and optionally included in PR mode if desired for a particular receiver topology).
- **TIA noise (PR mode):** implemented using both an *input-referred current noise density* $A/\sqrt{\text{Hz}}$ and an *output-referred voltage noise density* $V/\sqrt{\text{Hz}}$, enabling realistic separation of noise sources and direct mapping from datasheet-like parameters.

These abstractions align with how high-speed PAM4 receiver front-ends are characterized in practice, namely through *input-referred noise densities*,

⁸⁴ H. Li, C. Hsu, J. Sharma, J. Jaussi, and G. Balamurugan, “A 100-Gb/s PAM-4 Optical Receiver With 2-Tap FFE and 2-Tap Direct-Feedback DFE in 28-nm CMOS,” IEEE Journal of Solid-State Circuits, vol. 57, pp. 44-53, 2022.

bandwidth, and OMA/BER performance, rather than through low-level transistor noise modeling, while remaining physically grounded and parameterizable⁸⁵.

For bandwidth limitation, the surrogate currently uses a *one-pole low-pass* representation (enabled/disabled separately for the PD junction stage and the TIA stage). This captures the correct first-order ISI mechanism (exponential memory tail) and provides a controlled path for later upgrades (e.g., measured transfer-function hooks or higher-order responses). The importance of accurate bandwidth modeling is reinforced by recent 100G-class receiver demonstrations that explicitly integrate traveling-wave TIAs, current-mode sampling, or stacking/co-packaging to push bandwidth while keeping noise acceptable, precisely because the PD-TIA electrical response directly shapes the eye at symbol rates where $f_{3\text{dB}}$ is a limiting factor⁸⁶.

(iii) From received waveform to eye height, Q-factor, and BER (how the surrogate uses the front-end output)

Once the front-end output waveform is generated, subsequent receiver processing, sampling, clock recovery/decision timing, equalization (if enabled), and slicing produce standard eye/BER metrics. In this thesis, we treat the sampled decision variable for PAM4 as *level-conditioned distributions*:

- k indexes the transmitted PAM level (for PAM4, $k \in \{1,2,3,4\}$).
- μ_k denotes the *conditional mean* of the sampled voltage given level k .
- σ_k denotes the *conditional standard deviation* (noise + residual ISI/jitter effects) of the sampled voltage given level k .

These quantities provide the direct bridge from the physical front-end model to link metrics:

⁸⁵ D. Patel, A. Sharif-Bakhtiar, and T. Carusone, “A 112-Gb/s —8.2-dBm Sensitivity 4-PAM Linear TIA in 16-nm CMOS With Co-Packaged Photodiodes.”, IEEE Journal of Solid-State Circuits, vol. 58, pp. 771-784, 2023.

⁸⁶ J. Declercq, S. Niu, J. Lambrecht, B. Moeneclaey, C. Wang, and X. Yin, “Analog Demultiplexing 128-GBd Optical Receiver Integrating a Traveling-Wave TIA and a Current-Mode Sampling Front-End.”, Journal of Lightwave Technology, vol. 43, pp. 6250-6257, 2025.

- **Eye height** between adjacent levels is governed by $\mu_{k+1} - \mu_k$, which is reduced by bandwidth-induced ISI and by any compression/nonlinear gain in upstream optical blocks.

- The **Q-factor (per eye opening)** is governed by the ratio of level spacing to the combined noise spreads (captured by the σ_k terms). In practice, this metric is used to relate measured eye openings to expected BER under Gaussian approximations.

- **BER vs OMA/OSNR** trends emerge naturally when the front-end is swept over gain/noise/bandwidth settings and the resulting μ_k, σ_k are recomputed.

This is also where the surrogate's value for ML becomes explicit: training data must reflect the *correct conditional statistics* of a bandwidth-limited, noise-limited, multi-level IM/DD front-end. Otherwise, an AI optimizer may learn non-transferable decision boundaries (e.g., by underestimating signal-dependent noise or by learning unrealistic eyes because noise does not scale with bandwidth).

(iv) Implementation notes in the digital surrogate (traceability to code parameters)

In the MATLAB surrogate, this block is implemented as “rx_photodetector” with an explicit stage tag “PD_frontend”. The implementation is intentionally *plug-and-play*:

- **Input contract:** “signal.current” carries either $P(t)$ or $E(t)$; “p.rx.PD.inputIsPower_W” selects which.

- **Mode selection:** “p.rx.PD.mode $\in$ {photodiode, photoreceiver}” toggles PD vs PR behavior.

- **Physical parameters:** responsivity $\mathcal{R}$, dark current I_{dark} , load R_{load} , and TIA gain G_{TIA} are required parameters (no hidden defaults).

- **Noise control:** shot and thermal noise are toggled explicitly; PR mode also supports input- and output-referred TIA noise densities.

- **Bandwidth control:** one-pole PD bandwidth and one-pole TIA bandwidth can be enabled independently, allowing controlled experiments on ISI versus noise accumulation.

- **Metadata:** the block stores diagnostic RMS values (signal-only, noise-only, output) and a lightweight SNR estimate inside “signal.meta.PD.PD_frontend,” enabling reproducible dataset labeling for ML workflows.

This design aligns with the thesis goal: a physics-based surrogate that is accurate enough for system interpretation yet modular and parametric enough to serve as a data vessel for AI-based optimization.

6.4 ADC Surrogate Model

In a high-speed IM/DD receiver chain, the ADC is the final analog-to-digital boundary that captures all upstream optical and electrical impairments into the sampled sequence, which will later be equalized, classified, and decoded. Therefore, an end-to-end digital model that aims to generate data for AI-based optimization should not treat the ADC as a perfectly ideal sampler followed by an abstract SNR penalty: finite resolution and quantization noise can significantly alter the behavior and even the apparent benefit of learned DSP blocks when they are trained or evaluated on synthetic data. This sensitivity is explicitly seen in learning-based compensation studies where DAC/ADC quantization noise is a key limiting impairment, so ignoring the ADC’s non-idealities risks creating training data that overestimate achievable gains and do not transfer smoothly to the lab⁸⁷. Furthermore, in bandwidth-limited IM/DD transceivers, the interaction between analog conditioning (including pre-emphasis/equalization) and effective noise can reshape the SNR environment perceived by the ADC, underscoring the importance of modeling the ADC as a primary component in the surrogate rather than a secondary element⁸⁸.

⁸⁷ Q. Li, B. Chen, Y. Lei, and Z. Liang, “On the performance and fiber nonlinearity modeling of learned digital back-propagation in the presence of DAC and ADC quantization noise”, *Optics Express*, vol. 33, no. 4, pp. 7962-7977, 2025.

⁸⁸ Z. Zhou, T. Odedeyi, B. Kelly, J. O’Carroll, R. Phelan, I. Darwazeh, and Z. Liu, “Impact of Analog and Digital Pre-Emphasis on the Signal-to-Noise Ratio of Bandwidth-Limited Optical Transceivers”, *IEEE Photonics Journal*, vol. 12, pp. 1-12, 2020.

6.4.1 Why Model the ADC in a Physics-Based Digital Surrogate

Why the ADC cannot be treated as a passive “digitizer.”

In the Tx→Rx chain of an IM/DD system (including PAM4 and BPAM), the ADC is more than a measurement device; it is a signal-transformation block with its own bandwidth, nonlinearity, and noise mechanisms. In many link-level simulations, the ADC is simplified to “ideal sampling + AWGN,” but this is often insufficient when the goal is surrogate-driven optimization, because the optimization algorithm (especially ML-based) can exploit or be misled by details of the observation channel that are absent in the real hardware. A clear example is learning-based compensation, where DAC/ADC quantization noise alters performance trends and inferred optimal operating points, meaning that an ADC-free surrogate can produce data that trains the model toward a solution that is fragile in deployment⁸⁹.

Finite resolution: quantization is a structured impairment, not just “extra noise.”

For a uniform N -bit quantizer with full-scale span V_{FS} , the step size is $\Delta = V_{\text{FS}}/2^N$. Under the common (but explicitly stated) assumption that the quantization error is approximately uniform over $[-\Delta/2, \Delta/2]$ and weakly correlated with the signal, the quantization-noise variance is

$$\sigma_q^2 \approx \frac{\Delta^2}{12}.$$

This term establishes a strict lower limit on the achievable electrical SNR at the ADC output, which can become predominant in bandwidth-limited receivers where equalization and/or pre-emphasis modify the spectrum (and thus the effective peak-to-average behavior) of the sampled waveform. In such systems, the signal processing used to recover high symbol rates can heighten sensitivity to

⁸⁹ Q. Li, B. Chen, Y. Lei, and Z. Liang, “On the performance and fiber nonlinearity modeling of learned digital back-propagation in the presence of DAC and ADC quantization noise”, *Optics Express*, vol. 33, no. 4, pp. 7962-7977, 2025.

quantization by redistributing energy toward frequency regions that are harder to represent accurately. Therefore, the ADC model must be included to properly capture the end-to-end SNR/BER trade-offs⁹⁰. This point is particularly relevant for BPAM: because BPAM relies on *accurate multi-level thresholding after direct detection*, any quantization-induced compression or effective noise-floor increase can shrink the vertical eye opening and shift optimum operating points (e.g., gain settings, decision thresholds, equalizer parameters) in ways that a simplified ADC model would miss⁹¹.

Static nonlinearity (INL/DNL): why “datasheet metrics” may be insufficient for BER/EVM prediction.

ADC static non-idealities are typically summarized by DNL/INL, but these metrics do not uniquely predict communication-facing distortion (e.g., EVM): different shapes of converter transfer error can produce meaningfully different distortion patterns even when INL/DNL summaries appear similar. For communication-focused modeling and optimization, it is important to represent static conversion errors in a way that maintains their relationship to error vectors and ultimately BER, especially for multilevel formats where decision boundaries are close together, and level spacings are uneven after the full analog front-end⁹². In an ML-optimization setting, this is important because the learning system can “overfit” to a surrogate that lacks the correct static distortion structure, which decreases transferability to real measurements and necessitates larger transfer-learning adjustments later⁹³.

⁹⁰ Z. Zhou, T. Odedeyi, B. Kelly, J. O’Carroll, R. Phelan, I. Darwazeh, and Z. Liu, “Impact of Analog and Digital Pre-Emphasis on the Signal-to-Noise Ratio of Bandwidth-Limited Optical Transceivers”, IEEE Photonics Journal, vol. 12, pp. 1-12, 2020.

⁹¹ *Ibidem*. Q. Li et al. “On the performance and fiber nonlinearity modeling of learned digital back-propagation in the presence of DAC and ADC quantization noise”

⁹² M. Nazarathy and I. Tomkos, “All-Analytic Statistical Modeling of Constellations in (Optical) Transmission Systems Driven by High-Speed Electronic Digital to Analog Converters Part I: DAC Mismatch Statistics, Metrics, Symmetries, Error Vector Magnitude”, Photonics, 11(8), 747, 2024.

⁹³ *Ibidem*.

Sampling-time uncertainty and interleaving: timing errors become amplitude errors.

At very high symbol rates, aperture/clock jitter converts waveform slope into amplitude noise; for a differentiable input $v(t)$, a first-order approximation gives $e_j(t) \approx \dot{v}(t)\Delta t$. This mechanism tightens clocking requirements when high-frequency content is present (a frequent case in equalized PAM links), and it motivates including timing uncertainty as a *distinct* ADC impairment rather than lumping it into AWGN. Jitter-aware receiver architectures and front-end design explicitly treat sampling uncertainty as a first-class limitation for high-data-rate ADC-based reception⁹⁴. In addition, practical high-speed ADCs are frequently time-interleaved; in that case, gain/offset/timing mismatch between sub-ADCs can create spurious tones and structured distortion that cannot be summarized by a single ENOB number and must be budgeted at the system level when targeting BER/EVM⁹⁵.

Why is this ADC modeling essential for end-to-end optimization and lab transfer learning?

Because the ADC sits at the interface between the physical waveform and the digital algorithms, it effectively defines the “observation channel” for AI-based optimization: the learned equalizer/detector is trained to invert the *measurement operator implemented by the surrogate*. By incorporating finite-bit quantization and non-ideal conversion behavior, the surrogate produces training data with impairment statistics closer to the lab, reducing the domain gap and making it more realistic to: (i) optimize operating points end-to-end (e.g., gain/bandwidth/noise allocations across SOA-PD-ADC), and (ii) apply *transfer learning* using a smaller amount of measured data to fine-tune the model for a specific experimental setup, rather than retraining from scratch. The need for this alignment is reinforced by

⁹⁴ Y. Zhu, J. Diaz, S. Kaile, I. Yi, T. Liu, S. Hoyos, and S. Palermo, “A Jitter-Robust 40 Gb/s ADC-Based Multicarrier Receiver Front-End With 4-GS/s Baseband Pipeline-SAR ADCs in 22-nm FinFET”, IEEE Journal of Solid-State Circuits, vol. 58, pp. 662-676, 2023.

⁹⁵ A. Buchwald, “High-speed time interleaved ADCs”, IEEE Communications Magazine, vol. 54, pp. 71-77, 2016.

studies showing that quantization noise can directly influence the behavior and conclusions of learning-based compensation methods⁹⁶.

6.4.2 Sampling and clocking constraints: timing jitter, bandwidth, and BER

In an ADC-based IM/DD receiver chain, the sampling stage is not merely a “discretization” step: it defines a set of degrees of freedom (sampling rate, sampling phase, and timing uncertainty) that directly constrain the achievable BER through their interaction with the analog front-end bandwidth and the waveform’s effective slope. In the digital surrogate, we therefore treat the ADC sampling operation as a controllable interface between the continuous-time electrical waveform delivered by the photoreceiver front end and the discrete-time DSP chain, so that end-to-end optimization can explicitly account for these constraints rather than lumping them into a single “lumped” noise penalty. This perspective becomes even more important for multilevel PAM and bipolar PAM variants, where eye openings are tighter, and the receiver’s tolerance to sampling errors is reduced⁹⁷.

Sampling as an operator, and why reconstruction matters in a surrogate

Let $x(t)$ denote the analog electrical waveform at the ADC input, after the photoreceiver front-end and any analog filtering. An ideal uniform sampler produces

$$x[n] = x(nT_s), \quad T_s = \frac{1}{F_s}.$$

However, *any* realistic treatment of sampling rate changes (e.g., evaluating alternative ADC rates) or timing jitter requires evaluating $x(t)$ at *non-integer* time instants. A surrogate, therefore, needs an explicit reconstruction/evaluation

⁹⁶ Q. Li, B. Chen, Y. Lei, and Z. Liang, “On the performance and fiber nonlinearity modeling of learned digital back-propagation in the presence of DAC and ADC quantization noise”, *Optics Express*, vol. 33, no. 4, pp. 7962-7977, 2025.

⁹⁷ P. Miqueu, F. Belvèze, J. Brossier, and L. Ros, “Parameters setting based on bit error probability formula for an amplitude modulated, baseband, communication system affected by timing jitter and frequency selective channel”, *AEU - International Journal of Electronics and Communications*, 178, art. 155288, 2024.

mechanism, i.e., a mathematical way to represent the continuous-time waveform between the discrete points available inside the simulator.

For a band-limited signal (or a signal made effectively band-limited by the receiver's analog bandwidth limitation), the classical reconstruction model is

$$\hat{x}(t) = \sum_{k=-\infty}^{\infty} x[k] \operatorname{sinc}\left(\frac{t - kT_s}{T_s}\right),$$

with the understanding that practical implementations replace the ideal sinc with a realizable interpolation filter. In a digital surrogate, this reconstruction approach is the simplest way to (i) resample to different ADC rates and (ii) evaluate samples at jittered times without creating nonphysical artifacts. It also naturally aligns with generalized-sampling perspectives used to analyze sampling under mismatch and timing errors in real ADC architectures⁹⁸.

Timing (aperture) jitter: from time uncertainty to amplitude noise

With aperture jitter, the actual sampling instants deviate from the nominal grid:

$$t_n = nT_s + \delta t_n,$$

where δt_n is commonly modeled as a zero-mean random process with rms (σ_t). The sampled values become

$$x_j[n] = x(t_n) = x(nT_s + \delta t_n).$$

For sufficiently small jitter (relative to the time – scale over which $x(t)$ changes), a first-order Taylor expansion yields the fundamental mechanism:

$$x_j[n] \approx x(nT_s) + \delta t_n x'(nT_s).$$

⁹⁸ S. Ghosh and B. Sahoo, “Closed-Form Expression for the Combined Effect of Offset, Gain, Timing, and Bandwidth Mismatch in Time-Interleaved ADCs Using Generalized Sampling”, IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 2007112, 2021.

Hence the sampling error is

$$e[n] \triangleq x_j[n] - x(nT_s) \approx \delta t_n x'(nT_s),$$

which shows that timing jitter becomes *additive amplitude noise*, whose strength is proportional to the waveform slope. This is precisely why bandwidth, pre-emphasis/equalization, and symbol rate matter: they shape $x'(t)$, and therefore the jitter penalty.

For a single-tone $x(t) = A \sin(2\pi f t)$, this mechanism leads to the widely used jitter-limited SNR relationship

$$\text{SNR}_{\text{jitter}} \approx -20 \log_{10} (2\pi f \sigma_t) \quad (\text{dB}),$$

capturing the key scaling: higher frequency content (or steeper waveforms) and larger jitter both reduce the effective SNR. In broadband PAM signals, the same principle applies, but the relevant “ f ” is replaced by the signal’s effective high-frequency content that survives the receive filtering and equalization⁹⁹.

Bandwidth-limited PAM and BER impact: why “clocking” cannot be separated from “channel”

In bandwidth-limited PAM receivers, sampling jitter does not act in isolation: it interacts with the frequency selectivity of the receive chain (front-end bandwidth and any analog/electrical filtering) and with the DSP equalizer that may reshape the spectrum. Consequently, the BER penalty depends jointly on (i) the jitter statistics, (ii) the receiver bandwidth/selectivity, and (iii) the effective noise and ISI after filtering, precisely the set of interactions that an end-to-end surrogate must capture to generate physically meaningful training data for ML-based optimization.

A useful system-level framing is to express BER (or BEP) as a function of σ_t , the receive bandwidth/selectivity parameters, and the noise level, so that timing

⁹⁹ K. Sozański, “Overview of Signal Processing Problems in Power Electronic Control Circuits”, *Energies*, 16(12), 4774, 2023.

uncertainty is translated into a BER-relevant impairment rather than being treated as a generic “implementation loss.” This explicit BER formulation under timing jitter and frequency selectivity is particularly relevant for BPAM/PAM4-type signaling, where the smaller vertical eye margins make the BER more sensitive to sampling perturbations¹⁰⁰.

Architectural nonidealities: generalized sampling and interleaving mismatch as additional “knobs”

In many high-speed ADC implementations (especially time-interleaved architectures), sampling errors are not limited to a single jitter parameter; offset, gain, timing skew, and bandwidth mismatch across sub-ADCs can jointly degrade SNDR and introduce distortion components that are not captured by a single ENOB. From a surrogate perspective, this motivates keeping sampling/clocking impairments modular: timing uncertainty, bandwidth mismatch, and static offsets/gains can be turned on/off or calibrated separately, enabling (i) sensitivity analysis and (ii) transfer learning to lab measurements, where different boards/instruments exhibit different dominant mismatch signatures. Closed-form formulations that combine these mismatch terms under generalized sampling provide a rigorous foundation for deciding which degrees of freedom are worth exposing in the surrogate for optimization and calibration workflows¹⁰¹.

Practical relevance to end-to-end optimization and transfer learning in the lab

Treating sampling/clocking as explicit degrees of freedom enables two concrete workflows:

1. **End-to-end optimization:** the surrogate can jointly search over receiver bandwidth/equalization choices and sampling/clocking constraints

¹⁰⁰ P. Miqueu, F. Belvèze, J. Brossier, and L. Ros, “Parameters setting based on bit error probability formula for an amplitude modulated, baseband, communication system affected by timing jitter and frequency selective channel”, *AEU - International Journal of Electronics and Communications*, 178, art. 155288, 2024.

¹⁰¹ S. Ghosh and B. Sahoo, “Closed-Form Expression for the Combined Effect of Offset, Gain, Timing, and Bandwidth Mismatch in Time-Interleaved ADCs Using Generalized Sampling”, *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 2007112, 2021.

(rate/phase/jitter budget) to minimize BER under realistic hardware constraints.

2. **Transfer learning/calibration:** the surrogate’s sampling/clocking parameters can be tuned to match measured eye/EVM/BER trends from a specific lab setup (e.g., a given oscilloscope/ADC front-end), so that the model becomes a reusable “digital equivalent” of that measurement chain and can generate training data consistent with the experiment. Architectures that explicitly target jitter filtering and BER optimization in high-speed PAM-4 transceivers underscore the practical importance of the coupling between clocking strategy and BER performance¹⁰².

6.4.3 Quantization and ADC static nonlinearity (INL/DNL)

In the proposed digital surrogate, the ADC is modeled as an explicit receiver block instead of an ideal “measurement device”. This choice is intentional: in multilevel IM/DD reception (PAM4 and, more importantly, bipolar workflows where level geometry around the origin becomes crucial), finite-resolution code mapping and static transfer errors alter the effective decision regions seen by the DSP and any ML-based equalizer or decision module. If the surrogate generates training data with an ADC model that is too idealized, the learned model may overfit to unrealistically clean level spacing and will need significantly more fine-tuning when applied to laboratory traces. Conversely, a hardware-aware ADC model allows end-to-end optimization to explore receiver-limited regimes (resolution, clipping, and code-dependent distortion) and makes transfer learning more data-efficient, because the surrogate already incorporates the main converter mechanisms¹⁰³.

Ideal uniform quantization model, step size, and ENOB view

¹⁰² H. Xu, M. Lai, X. Zheng, Z. Wang, J. Xu, S. Li, F. Lv, M. Liu, W. Li, Z. Wen, X. Guo, X. Wang, Z. Jin, and X. Liu, “A 112-Gb/s PAM-4 Retimer Transceiver With Jitter-Filtering Clocking Scheme and BER Optimization Technique in 28-nm CMOS”, IEEE Journal of Solid-State Circuits, vol. 60, pp. 2305-2318, 2025.

¹⁰³ A. Verreault, P. Cicek, and A. Robichaud, “Oversampling ADC: A Review of Recent Design Trends”, IEEE Access, vol. 12, pp. 121753-121779, 2024.

Let $v[n]$ denote the sampled electrical waveform at the ADC input (after anti-alias filtering and sampling). The surrogate implements a uniform N -bit quantizer by first defining a full-scale span V_{FS} and the associated least significant bit (LSB). Importantly, the ADC and DAC sides do not use the same “LSB convention” in our implementation:

- **DAC quantizer (TX side):** uses an *endpoint-inclusive* step $Q = V_{\text{FS}}/(2^N - 1)$, which produces a symmetric mid-rise ladder with no exact zero reconstruction level (consistent with the finite even number of levels).
- **ADC quantizer (RX side):** uses an LSB $\Delta = V_{\text{FS}}/2^N$ and reconstructs levels as

$$\hat{v}[n] = V_{\min} + c[n]\Delta,$$

with integer code $c[n] \in \{0, \dots, 2^N - 1\}$. With $(V_{\min} = -V_{\text{FS}}/2)$, this construction contains an exact zero reconstruction level at $c = 2^{N-1}$, i.e., a mid-tread behavior around the origin.

Under the standard “busy input” approximation, $e_q[n] = v[n] - \hat{v}[n]$ would be the quantization error, treated as approximately uniform on $(-\Delta/2, \Delta/2)$, yielding the classical variance

$$\sigma_q^2 \approx \frac{\Delta^2}{12}.$$

This establishes a baseline “resolution-limited” noise floor and serves as the reference against which ENOB/SNDR are interpreted when the ADC employs oversampling or noise-shaping (which redistributes quantization noise out of band)¹⁰⁴. In receiver optimization studies, this is convenient: a single parameter (effective resolution/ENOB) can sweep the surrogate between quantization-limited and channel/noise-limited regimes while keeping the rest of the chain unchanged¹⁰⁵.

¹⁰⁴ *Ibidem*.

¹⁰⁵ C. Rodenbeck, M. Martinez, J. Beun, J. Silva-Martínez, A. Karsilayan, and R. Liechty, “*When Less Is More ... Few Bit ADCs in RF Systems*”, IEEE Access, vol. 7, pp. 12035-12046, 2019.

Mid-rise vs mid-tread quantizers and why the distinction matters for PAM-level geometry

A critical modeling nuance is whether the uniform quantizer is implemented as *mid-rise* or *mid-tread*, because this changes the reconstruction levels and the behavior near zero, and therefore changes the *geometry* of small-signal errors around the most sensitive decision region.

- **Mid-tread quantizer (includes a zero reconstruction level):** thresholds are at $(k + \frac{1}{2})\Delta$ and reconstruction levels are at $k\Delta$. The central cell maps to exactly zero, producing a “dead-zone” of width Δ :

$$Q_{\text{midtread}}(v) = \Delta \cdot \text{clip}\left(\text{round}\left(\frac{v}{\Delta}\right)\right).$$

This “dead-zone” structure is explicitly exploited in dead-zone quantization designs, where robustness around zero is a desired property.¹⁰⁶

- **Mid-rise quantizer (no zero reconstruction level):** thresholds are at $k\Delta$ and reconstruction levels are at $(k + \frac{1}{2})\Delta$, so the output never equals exactly zero:

$$Q_{\text{midrise}}(v) = \Delta \cdot \text{clip}\left(\left\lfloor \frac{v}{\Delta} \right\rfloor + \frac{1}{2}\right).$$

Mid-rise characteristics are also treated analytically in few-bit quantization settings (e.g., mid-rise TDC/quantizer models), where the central “no-zero-level” behavior is part of the modeled transfer¹⁰⁷.

Why this matters in IM/DD PAM and bPAM contexts.

¹⁰⁶ S. Li, H. Li, W. Dai, C. Li, J. Zou, and H. Xiong, “*Learned Progressive Image Compression With Dead-Zone Quantizers*”, IEEE Transactions on Circuits and Systems for Video Technology, vol. 33, pp. 2962-2978, 2023.

¹⁰⁷ X. Wang and M. Kennedy, “*Linearized Analysis and Quantization Error Minimization for Mid-Rise TDCs: A Tutorial*”, IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 72, pp. 2986-2997, 2025.

In multilevel PAM reception, the most significant errors often arise from disturbances to the relative spacing of levels and to the statistics of samples near decision boundaries. A mid-tread ADC produces an exact zero code and a dead zone around the origin; a mid-rise ladder removes the zero reconstruction level and forces the smallest magnitudes to map to $\pm\Delta/20$. In practice, these two behaviors cause measurably different small-signal distortion patterns and therefore alter EVM/BER sensitivity when the waveform distribution is concentrated near the center (which is common after equalization in bandwidth-limited links, and can be especially significant in bipolar processing chains).

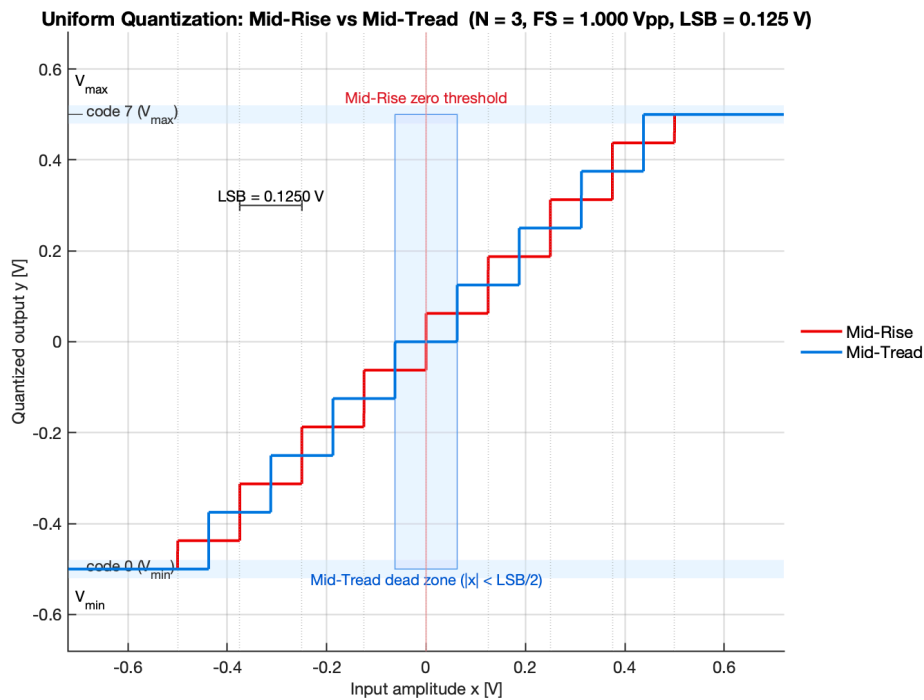


Figure 6.x Mid-rise vs Mid-Tread step sizes in uniform quantization: thresholds, reconstruction levels, and the zero-region behavior

Implementation note (surrogate): the Rx quantizer is implemented as *mid-tread* by construction ($\Delta = V_{FS}/2^N$, explicit integer code c , and a reconstruction level exactly at zero), whereas the Tx quantizer uses a *mid-rise* ladder (endpoint-inclusive step $V_{FS}/(2^N - 1)$). This deliberate asymmetry prevents “accidental agreement” between the transmitter and receiver quantizers and makes the surrogate more informative for end-to-end optimization studies in which the receiver ADC is a bottleneck.

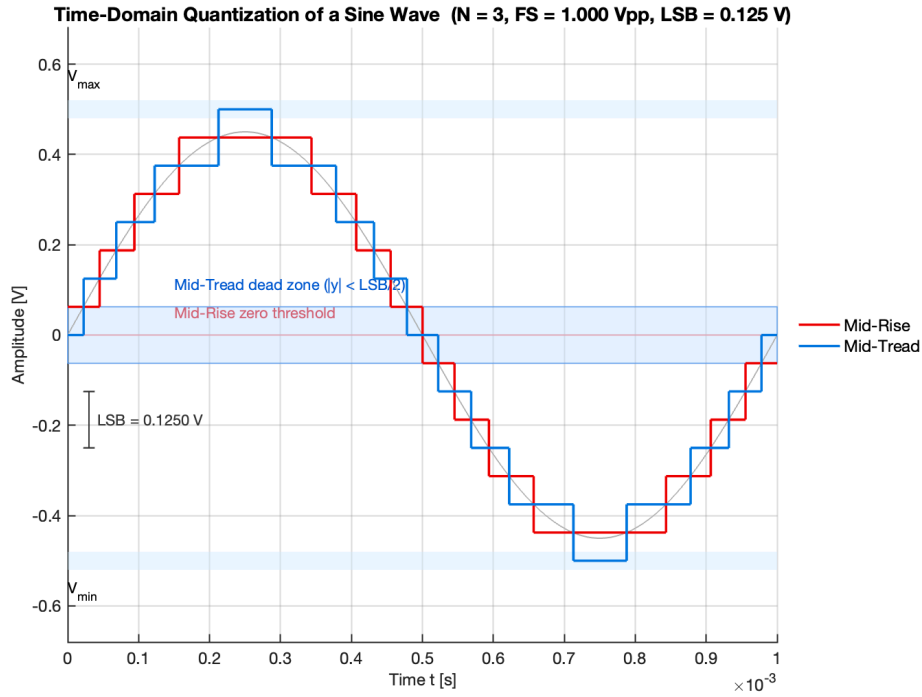


Figure 6.xx Mid-rise vs mid-tread: Time domain Quantization of a sine wave

ADC static nonlinearity: INL/DNL as threshold perturbations (injection point in the surrogate)

Beyond ideal uniform quantization, the ADC is affected by **static transfer errors** that are naturally expressed as perturbations of code transition thresholds:

- Let the ideal transition thresholds be $T_k = V_{min} + k\Delta, k = 0, \dots, 2^N$.
- Let the perturbed thresholds be $T'_k = T_k + \epsilon_k$, with a *monotonic constraint* $T'_{k+1} > T'_k$.

The standard definitions can then be expressed as

$$DNL_k = \frac{T'_{k+1} - T'_k}{\Delta} - 1, \quad INL_k = \frac{T'_k - T_k}{\Delta},$$

where DNL captures *step-width* errors, and INL captures *cumulative* deviation from the ideal straight-line transfer.

How INL/DNL is injected in our ADC quantizer

In the surrogate, INL/DNL is applied *before* code assignment by operating directly on the threshold vector T :

1. construct ideal T_k from (V_{min}, Δ, N)
2. apply an INL/DNL profile to produce a monotonic T'_k ;
3. quantize each sample by locating the interval $[T'_k, T'_{k+1})$ that contains $v[n]$ (threshold-based binning), yielding an integer code $c[n]$;
4. reconstruct $\hat{v}[n] = V_{min} + c[n]\Delta$

This threshold-perturbation approach is particularly suitable for multilevel PAM because it directly models decision-boundary displacement (level-dependent slicing bias), which can increase EVM and elevate BER even when the additive noise floor is unchanged. When such structured distortion is present, BER behavior may deviate from what is predicted by SNR-only arguments unless the distortion is reflected in the effective error statistics (e.g., via EVM-based assessment)¹⁰⁸.

During the development of this surrogate, a device-specific measured INL/DNL dataset for the exact laboratory ADC was not available; therefore, the surrogate currently uses a controlled artificial profile family (bounded DNL with correlated INL and enforced monotonicity). This is not merely a stopgap: it enables robustness training by exposing ML equalizers/decision modules to plausible code-dependent distortions. It later provides a clean transfer-learning pathway, in which measured threshold maps can replace the synthetic profile, and the ML model can be fine-tuned on a smaller set of lab traces. In related converter contexts (e.g., interleaved ADC mismatch), it is common to explicitly model and then calibrate such static/mismatch errors rather than collapsing everything into a single ENOB number¹⁰⁹.

¹⁰⁸ H. Qin and X. Xiao, “Effects of fiber nonlinearity on error vector magnitude and bit error ratio for advanced modulation formats”, *Optical Engineering*, vol. 57, 2018.

¹⁰⁹ C. Schmidt, J. Figueroa, J. Cousseau, and A. Tonello, “Pilot-Based TI-ADC Mismatch Error Calibration for IR-UWB Receivers”, *IEEE Access*, vol. 7, pp. 74340-74350, 2019.

6.4.4 Anti-Aliasing Filters (AAF) and ADC Noise-Separation Strategy

In an end-to-end IM/DD digital surrogate, the ADC is not only a “quantizer,” but also the final analog-digital boundary where (i) bandwidth selection, (ii) sampling-time uncertainty, (iii) code non-idealities, and (iv) an ADC-internal noise floor collectively determine the discrete-time waveform that will be equalized and detected. A key reason for explicitly modeling the AAF and including a separate ADC-noise term is that optimization and transfer learning become more effectively formulated: we can tune and identify front-end/channel impairments separately from converter-internal impairments, instead of forcing the learning loop to explain everything with a single “effective noise” term¹¹⁰.

AAF as an LTI Bandwidth-Selection Operator

Let the continuous-time (or high-rate discrete-time) input to the ADC be $x(t)$. The anti-aliasing filter is modeled as an LTI system with impulse response $h_{\text{AAF}}(t)$ and frequency response $H_{\text{AAF}}(f)$. The filtered waveform is

$$x_{\text{AAF}}(t) = (x * h_{\text{AAF}})(t), \quad X_{\text{AAF}}(f) = X(f)H_{\text{AAF}}(f).$$

Sampling at (F_s) then yields $x_{\text{ADC}}[n] = x_{\text{AAF}}(nT_s)$ with $T_s = 1/F_s$. The AAF objective is to attenuate $X(f)$ for $|f| > F_s/2$ so that spectral replicas do not fold into baseband. In high-speed converters, this AAF is often implemented using compact, high-order responses (or RF structures) precisely because out-of-band attenuation must be achieved over tens of GHz¹¹¹.

Magnitude vs. group-delay trade-off.

A key reason we treat the AAF explicitly is that filter families differ not only in roll-off but also in phase linearity. In time-domain sensitive links (PAM and equalized IM/DD), *group-delay variation* across the signal band can distort

¹¹⁰ A. Verreault, P. Cicek, and A. Robichaud, “Oversampling ADC: A Review of Recent Design Trends,” IEEE Access, vol. 12, 2024.

¹¹¹ A. Lorenz, C. Matthus, and F. Ellinger, “Microstrip-Based Anti-Aliasing Filters for Beyond-50 GHz-Bandwidth Mix-Filter-Sample Analog-to-Digital-Converters,” IEEE Access, vol. 13, 2025.

transitions and interact with equalization and clock recovery. This motivates considering responses that constrain group delay (Bessel-like or low-group-delay IIR designs), even when their magnitude roll-off is gentler than a Butterworth of comparable order¹¹².

Noise-Equivalent Bandwidth (NEB): the AAF’s “Noise Gain” in Frequency

When the input carries (approximately) white voltage noise with two-sided PSD $S_x(f) = N_0/2$, the AAF shapes the integrated in-band variance. The standard quantity that captures this is the *noise-equivalent bandwidth*:

$$B_n \triangleq \frac{1}{|H_{\text{AAF}}(0)|^2} \int_0^\infty |H_{\text{AAF}}(f)|^2 df,$$

so that the output noise variance (for unit DC gain) is approximately

$$\sigma_{\text{AAF}}^2 \approx N_0 B_n.$$

This matters directly for BER because any “bandwidth extension” strategy that increases high-frequency gain (peaking/equalization) may also raise B_n , i.e., it can gather more noise power even if the (-3) dB bandwidth appears attractive. In wideband filtering-ADC-style thinking, it is common to relate architectures to an equivalent Butterworth-like response to connect filter shape to integrated noise at the system level¹¹³.

How the Digital Surrogate Implements the AAF (and Why)

Our implementation applies the AAF **before** the ADC sampling stage, which is the physically consistent order for alias control: filtering must precede any rate

¹¹² H. Yi, F. Chang, P. Tan, B. Huang, Y. Wu, Z. Xu, B. (et al.), “Design of infinite impulse response maximally flat stable digital filter with low group delay,” Scientific Reports, vol. 15, 2025.

¹¹³ S. Pavan, S. Manivannan, and N. Basavaraj, “Analysis and Design of Wideband Filtering ADCs Using Continuous-Time Pipelining,” IEEE Journal of Solid-State Circuits, vol. 59, 2024.

reduction. Specifically, the ADC parent function calls the AAF function, which then invokes the sampler. Inside the ADC AAF function, the AAF is implemented as a configurable IIR filter (Butterworth / Chebyshev / elliptic; and an analog-prototype Bessel option via a bilinear transform), optionally followed by a *notch cascade* (implemented as one or more biquads) to suppress narrowband interferers or structured tones. This gives the surrogate a clean “hook” for measured-transfer-function matching later, while keeping a compact parametric form for optimization.

A key nuance we maintain is that “AAF bandwidth” is not a single value in the code: the design can be manual (explicit pass/stop edges) or auto/semi-optimized (derive edges from estimated occupied bandwidth, then calibrate the achieved -3 dB point). This is intentional: it allows an optimizer to adjust only a small set of degrees of freedom (e.g., a bandwidth cap and an order), while the function manages realizability and stability constraints internally.

ADC Noise Classes and Why We Separate Them

For end-to-end learning (and especially sim-to-lab transfer), it is useful to distinguish *where* noise is injected, because it changes what the model can learn and how parameters can be identified:

1. **Front-end/channel noise (pre-ADC):** noise already present at the ADC input (e.g., from PD+TIA, ASE/RIN mappings, electrical interference). This noise is *shaped by the AAF*; hence, its contribution is governed by $H_{\text{AAF}}(f)$ and B_n .
2. **ADC internal analog noise (pre-quantization):** sampling switch thermal noise, comparator/amplifier noise, reference noise coupling, etc. In a detailed model, much of this is naturally treated as an *input-referred* noise term added *before* quantization, and it may also be spectrally shaped depending on the architecture. Oversampling and VCO-based/noise-shaping ADC literature explicitly separates internal noise-shaping and quantization mechanisms from

external noise sources, which is exactly the modeling philosophy we want for ML-ready surrogates¹¹⁴.

3. **Quantization noise/code uncertainty (at quantization):** the deterministic mapping plus any modeled threshold non-uniformity (INL/DNL) defines an effective distortion/noise floor even when the additive noise is small.

4. **Digital/output-referred noise (post-quantization):** numerical noise, additional measurement uncertainty, or a compact “effective ENOB loss” term.

Currently, our ADC noise injection is achieved by adding a white Gaussian sequence with either (i) a target SNR relative to a reference amplitude or (ii) a specified sigma directly. In the current process, this AWGN is applied after quantization, making it an output-referred additive term. This is a valid first-order approach when we need a simple, adjustable floor for system sweeps and ML training baselines.

Why separation is still worth pursuing

If we lump everything into a single AWGN knob, the learning loop can fit the BER, but it cannot reliably explain *why it* changes (aliasing vs. quantization vs. front-end noise vs. ADC-internal noise). By keeping at least two independent noise “families” (pre-AAF/front-end vs. ADC-internal), we can:

- Calibrate the surrogate to lab measurements by tuning a smaller, interpretable subset of parameters (transfer learning/domain adaptation),
- Expose cleaner latent factors to ML models (better generalization and robustness),
- Avoid compensating one impairment with another in optimization (e.g., “learning” a wrong equalizer because the simulator noise is placed at the wrong node)¹¹⁵.

¹¹⁴ B. Saux, J. Borgmans, J. Raman, and P. Rombouts, “A 3.5 GS/s 1-1 MASH VCO ADC With Second-Order Noise Shaping,” IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 72, 2024.

¹¹⁵ T. Wu, C. Ho, and M. Chen, “A Flash-Based Non-Uniform Sampling ADC With Hybrid Quantization Enabling Digital Anti-Aliasing Filter,” IEEE Journal of Solid-State Circuits, vol. 52, 2017.

A practical next step, without increasing complexity too much, is to split the current AWGN into (a) *input-referred pre-quantization noise* (added before the quantizer) and (b) a smaller *post-quantization output-referred* term reserved for digital measurement uncertainty. This preserves identifiability while remaining computationally light.

Chapter 7

Digital Receiver Section: DSP and Machine Learning

This chapter examines the digital receiver side of short-reach IM-DD links, focusing on PAM4/BPAM systems in which bandwidth limits, chromatic-dispersion-related impairments, and practical device nonlinearities render the end-to-end channel a finite-memory (often nonlinear) ISI challenge. Starting from a discrete-time signal model, the chapter explains why receiver DSP is crucial in modern IM-DD interconnects: it restores eye opening, stabilizes decision thresholds, and provides reliable hard/soft decisions amidst realistic noise and implementation constraints.

Building on this foundation, the chapter walks through the receiver pipeline from front-end digital processing (e.g., resampling/time-axis consistency and basic conditioning) to the core topic of equalization. It first consolidates the classical baseline, linear FFE/DFE viewpoints (ZF/MMSE), adaptive methods (LMS/RLS and decision-directed variants), and nonlinear/sequence approaches (Volterra-type structures and MLSE/BCJR references). It then cleanly separates machine-learning equalizers as a complementary family aimed at imperfect modeling, nonlinearities, and long memory. The chapter concludes by extending the discussion to BPAM and presenting an experimental demonstration of CNN-based equalization, linking the algorithmic narrative to practical validation and receiver-side robustness considerations.

Importantly, this receiver-focused discussion also serves as the final link in the digital surrogate framework developed throughout the thesis. The surrogate's transmitter blocks, electro-optic front-ends, and channel models define not only the waveform impairments observed at the receiver but also the precise DSP requirements, operating margins, and sensitivity to drift and hardware variability. In this sense, Chapter 7 'uses' the surrogate's physically grounded end-to-end outputs (including memory and nonlinearity signatures and their metadata) and

translates them into actionable receiver strategies, whether classical or ML-based, thereby completing the loop from modeled physical impairments to optimized digital recovery and, ultimately, to surrogate-driven design and validation of robust IM-DD systems.

7.1 Receiver-DSP Overview for IM-DD PAM4

Short-reach IM/DD PAM4 links depend on receiver-side DSP to address bandwidth limitations and the resulting ISI, while keeping costs and power consumption lower than coherent solutions. In modern post-200G/lane designs, the DSP challenge is typically viewed as detection over a sampled channel with limited memory, with additional penalties from device transfer characteristics and square-law detection. The receiver architecture is chosen to balance performance with implementation complexity¹.

7.1.1 Discrete-time received-signal model (sampling, channel memory, noise, nonlinearities)

Let $a[k] \in \{\pm 1, \pm 3\}$ denote the PAM4 symbol sequence (normalized levels), transmitted with symbol period T . After pulse shaping and the cascaded electrical/optical front-end responses, the *received optical intensity* at the photodiode can be viewed (at the system level) as a bandwidth-limited waveform corrupted by noise and distortions. A convenient receiver abstraction is to express the ADC input as

$$u(t) = (\mathcal{H}\{x(t)\})(t) + v(t), \quad x(t) = \sum_k a[k] p(t - kT),$$

where $\mathcal{H}(\cdot)$ denotes the effective continuous-time response of the end-to-end IM/DD chain (including bandwidth limits and dispersion-induced power-fading effects), and $v(t)$ aggregates noise contributions at the receiver front end. This

¹ D. Che, X. Chen, “Modulation Format and Digital Signal Processing for IM-DD Optics at Post-200G Era,” *Journal of Lightwave Technology*, vol. 42 (2), pp. 588-605, 2024.

“effective channel with memory” abstraction is the starting point for most DSP and equalizer designs in short-reach IM/DD PAM systems².

After sampling at $t = nT_s$ and ADC quantization $Q(\cdot)$, the discrete-time observation is

$$r[n] = Q(u(nT_s)) = Q\left(\sum_{\ell=0}^{L_h-1} h[\ell] s[n-\ell] + w[n]\right)$$

where $s[n]$ is the sampled pulse-shaped symbol waveform, $(h[\ell])$ is a finite-memory discrete-time representation of the effective channel, and $w[n]$ denotes sampled noise. For complexity control, many receiver designs deliberately approximate this as a short-memory ISI channel, enabling reduced-state detection and low-complexity equalization while still capturing the dominant bandwidth-limited ISI behavior relevant to PAM4 links³.

Square-law detection and device nonlinearities cause deviations from a purely linear ISI + AWGN model. In practice, nonlinear and pattern-dependent effects can be managed either by (i) maintaining a linear ISI model but increasing memory and employing sequence detection, or (ii) enhancing the model with structured nonlinear terms and applying nonlinear equalization or detection. A typical example is the use of MLSE-style detection in bandwidth-limited PAM4 links, where the receiver explicitly leverages finite channel memory to make better decisions compared to memoryless slicing⁴. When nonlinear distortion becomes significant,

² T. Wettlin, S. Calabrò, T. Rahman, J. Wei, N. Stojanović, S. Pachnicke, “*DSP for High-Speed Short-Reach IM/DD Systems Using PAM*,” *Journal of Lightwave Technology*, vol. 38 (24), pp. 6771-6778, 2020.

³ Y. Yu, Y. Che, T. Bo, D. Kim, H. Kim, “*Reduced-state MLSE for an IM/DD system using PAM modulation*,” *Optics Express*, vol. 28(26), pp. 38505-38515, 2020.

⁴ Z. Chen, J. Nie, S. Zhang, Q. Yang, X. Dai, L. Deng, M. Cheng, D. Liu, “*56-Gb/s/λ C-band DSB IM/DD PAM-4 40-km SSMF transmission employing a multiplier-free MLSE equalizer*,” *Optics Express*, vol. 30(7), pp. 11275-11287, 2022.

joint pulse shaping and low-complexity nonlinear equalization are often used to reduce residual impairments that linear equalizers cannot eliminate⁵.

In addition to additive noise, IM/DD PAM links can suffer from *multipath interference (MPI)*, in which delayed, reflected replicas create intensity fluctuations with level-dependent statistics that may deviate from simple Gaussian assumptions. Such impairments motivate specialized mitigation and/or detection strategies beyond standard linear equalization⁶.

7.1.2 Canonical DSP chain

A typical offline DSP process for IM/DD PAM4 starts with data capture and resampling: the waveform from the oscilloscope or ADC is first converted to a convenient sampling rate (often 1–2 samples per symbol) and aligned with the symbol grid through coarse synchronization and timing adjustments. This step creates the discrete-time signal on which further equalization and detection are based⁷.

The next stage typically includes *front-end conditioning*, such as low-pass (or matched) filtering and basic normalization (e.g., DC removal and scaling) to stabilize downstream adaptation and detection. In bandwidth-limited PAM4 links, these operations are often framed as part of practical channel/noise conditioning before equalization and soft/hard detection⁸.

Equalization is then applied gradually. A common baseline is a linear fractionally-spaced FFE to reduce bandwidth-limited ISI and improve robustness

⁵ X. Wu, J. Zhang, A. Lau, C. Lu, “C-band 120-Gb/s PAM-4 transmissions over a 100-km dispersion-uncompensated SSMF using joint combined pulse shaping and low-complexity nonlinear equalization,” *Optics Letters*, vol. 47(19), pp. 5144–5147, 2022.

⁶ Y. Li, Z. Wei, L. Liu, H. Fu, C. Yu, “Mitigated multipath interference in PAM4 IM/DD optical links using instantaneous mean intensity addressing,” *Optics Express*, Vol. 33, Issue 9, pp. 18855–18865, 2025.

⁷ L. Zhou, X. Fang, Y. Zhu, J. Zhu, Y. Li, Z. Chen, X. Nie, C. Yang, X. Zhou, Z. He, L. Wang, K. Li, F. Zhang, “Up to 182 GBd PAM-4 optical interconnects eliminating receiver-side equalization DSP,” *Optics Letters*, Vol. 50(7), pp. 2263–2266, 2025.

⁸ T. Wettlin, S. Calabrò, T. Rahman, J. Wei, N. Stojanović, S. Pachnicke, “DSP for High-Speed Short-Reach IM/DD Systems Using PAM,” *Journal of Lightwave Technology*, vol. 38 (24), pp. 6771–6778, 2020.

against sampling-phase errors when operating above 1 sample per symbol. In more difficult ISI scenarios, receivers may use DFE-type structures; although effective, DFE-based solutions are known to cause burst-error behavior, which in turn encourages soft-decision detection strategies designed specifically for DFE error propagation⁹.

For performance closer to sequence-optimal detection with finite memory, many IM/DD PAM4 receivers use MLSE (or reduced-state MLSE) variants, often implemented with carefully designed branch metrics and memory reduction methods to keep complexity manageable. This approach is especially natural when the link can be accurately approximated by a short-memory channel model¹⁰.

When nonlinearities dominate (from transmitter/receiver devices, bandwidth–dispersion interactions, or other combined distortions), the typical chain is extended with nonlinear equalization techniques (e.g., low-complexity nonlinear equalizers combined with pulse shaping), which address residual penalties that stay after linear ISI compensation¹¹.

Specialized blocks may also be introduced for specific impairment classes. For example, *MPI-aware mitigation* can be added when reflections introduce significant intensity noise or level-dependent disturbances, complementing the baseline equalizer/detector stack¹².

Finally, detection and performance metrics are calculated. Hard decisions allow for BER evaluation, while soft-information-based processing supports FEC-relevant metrics. In high-speed IM/DD PAM systems, reporting typically includes

⁹ K. Wu, G. Liga, J. Riani, A. Alvarado, “Low-Complexity Soft-Decision Detection for Combating DFE Burst Errors in IM/DD Links,” *Journal of Lightwave Technology*, vol. 42(5), pp. 1395-1408, 2023

¹⁰ Y. Yu, Y. Che, T. Bo, D. Kim, H. Kim, “Reduced-state MLSE for an IM/DD system using PAM modulation,” *Optics Express*, vol. 28(26), pp. 38505-38515, 2020.

¹¹ X. Wu, J. Zhang, A. Lau, C. Lu, “C-band 120-Gb/s PAM-4 transmissions over a 100-km dispersion-uncompensated SSMF using joint combined pulse shaping and low-complexity nonlinear equalization,” *Optics Letters*, vol. 47(19), pp. 5144-5147, 2022.

¹² W. Cui, J. Huo, J. Zhu, J. Tao, P. Qin, X. Zhang, H. Bai, “Machine Learning-Assisted Mitigation of Optical Multipath Interference in PAM4 IM-DD Transmission Systems,” *Photonics*, 12(4), 310, 2025.

BER and may also encompass information-theoretic measures such as GMI/NGMI when advanced DSP and probabilistic shaping are used to achieve high net bit rates within FEC constraints¹³.

7.1.3 Where DSP interfaces the digital surrogate blocks

From a system-integration perspective, the DSP chain above requires that the upstream simulator (or digital surrogate) provide the correct “receiver-facing” observables and metadata. At minimum, DSP realism depends on supplying waveform samples at a specific sampling rate (and resampling policy), the effective discrete-time channel response (which captures bandwidth limits and any dispersion-induced power fading), and noise statistics that match the physical operating point. These quantities influence equalizer memory, detector design, and the validity of the assumed discrete-time model¹⁴.

Equally important are the impairment models that challenge idealized assumptions. If the surrogate is used to evaluate nonlinear equalizers or nonlinear-aware detection, it must produce waveforms that maintain the relevant nonlinear signatures (e.g., device transfer curves and composite distortions that survive sampling). This explains why recent work has examined data-driven surrogates that learn experimental IM/DD behavior and can generate realistic received waveforms for DSP evaluation.¹⁵

Hardware realism also matters whenever the DSP is evaluated under implementation constraints. In such cases, the surrogate should include quantization and clipping effects (DAC/ADC resolution limits) and be able to output the ground-

¹³ A. Yan, G. Li, S. Xing, W. Shen, Y. Hu, A. Sun, J. Shi, Z. Li, C. Shen, N. Chi, J. Zhang, “*Net-432-Gb/s/λ PS-PAM transmission based on advanced DSP and single DAC for 400 G/λ IM/DD links*,” *Optics Letters*, Vol. 49, Issue 7, pp. 1656-1659, 2024.

¹⁴ T. Wettlin, S. Calabrò, T. Rahman, J. Wei, N. Stojanović, S. Pachnicke, “*DSP for High-Speed Short-Reach IM/DD Systems Using PAM*,” *Journal of Lightwave Technology*, vol. 38 (24), pp. 6771-6778, 2020.

¹⁵ T. Zhang, W. Liu, Z. Zhong, H. Li, Y. Gong, Q. He, M. Luo, S. Liu, “*A TCN Based Data-Driven Method Combined With TL for Modeling in Experimental IM-DD System*,” *IEEE Photonics Technology Letters*, vol. 37(2), pp. 105 – 108, 2025.

truth labels (transmitted symbols/bits) needed for supervised DSP/ML development and reliable metric calculation. This is especially critical in shaped or advanced-format demonstrations, where accurate priors and labels are required to compute FEC-relevant performance measures consistently¹⁶.

Within this thesis, the developed digital surrogate offers a systematic way to connect physical-layer blocks to receiver DSP by enforcing (i) standardized signal-domain interfaces, (ii) controllable impairment injection with scientifically meaningful parameters, and (iii) metadata generation aligned with DSP needs (sampling rate, timing offset/jitter models, quantization parameters, and ground truth). This allows for quick, repeatable receiver studies that would be costly to perform solely in the laboratory: one can stress-test equalizers and detectors across bandwidth limits, noise levels, and distortion regimes; compare linear, sequence, nonlinear, and ML equalizers under the same conditions; and run controlled ablations to identify which physical non-idealities most influence the observed BER/EVM/GMI trends. Additionally, by calibrating surrogate parameters to a limited set of measured traces, the same framework can predict performance and help guide co-design decisions (e.g., how much transmitter-side conditioning is needed to reduce receiver DSP), complementing emerging efforts to shift complexity away from receiver equalization when feasible¹⁷.

7.2 Front-End Digital Processing (Common PAM4 Practice)

The front-end DSP in short-reach IM/DD PAM4 receivers is responsible for converting a high-speed oscilloscope or ADC capture into a timing-aligned, symbol-synchronous discrete-time sequence that can be equalized and detected.

¹⁶ K. Wang, L. Zhang, C. Yu, C. Wang, Y. Chen, J. Yu, “FPGA-based 4×29.4912 Gbit/s PS-PAM4 signal transmission with a low-complexity probabilistic shaping scheme,” *Optics Letters*, vol. 48(6), pp. 1514-1517, 2020.

¹⁷ L. Zhou, X. Fang, Y. Zhu, J. Zhu, Y. Li, Z. Chen, X. Nie, C. Yang, X. Zhou, Z. He, L. Wang, K. Li, F. Zhang, “Up to 182 GBd PAM-4 optical interconnects eliminating receiver-side equalization DSP,” *Optics Letters*, Vol. 50(7), pp. 2263-2266, 2025.

Although the physical link involves square-law photodetection and bandwidth-limited optoelectronics, the front-end is usually designed to (i) establish a well-defined time axis and sampling rate, (ii) lock the sampling phase to the symbol clock (often through fractionally spaced processing), (iii) correct for slow offsets and normalize amplitude so adaptive algorithms operate in a stable numerical environment, and (iv) implement practical filtering that balances noise reduction with additional ISI before equalization¹⁸.

7.2.1 Resampling / sample-rate alignment and time-axis consistency

Let the captured real-valued waveform from the receiver front-end be $r_{\text{acq}}[n]$, sampled at $F_{s,\text{acq}}$ (oscilloscope/ADC clock). In offline PAM4 experiments, $F_{s,\text{acq}}$ is generally chosen to be substantially larger than the symbol rate $R_s = 1/T$ to provide oversampling margin for subsequent timing recovery and fractionally spaced equalization. A common engineering goal is to convert the acquisition-rate record into a working-rate sequence with a convenient samples-per-symbol (Sa/sym), most often (2) Sa/sym (i.e., $T/2$ spacing), because this simultaneously supports timing recovery and robust linear equalization under sampling-phase errors¹⁹.

To maintain time-axis consistency, the resampling should be expressed explicitly as a *time mapping* rather than as an index-only operation. Defining the acquisition time grid $t_n = n/F_{s,\text{acq}}$, resampling constructs a new uniform grid $t'_m = m/F_{s,\text{work}}$, where $F_{s,\text{work}} = MR_s$ with $M \in \{1,2\}$ typically, and then interpolates $r_{\text{acq}}(t)$ onto t'_m to produce $r[m] \approx r_{\text{acq}}(t'_m)$. In practice, the resampling / interpolation is tightly coupled to timing recovery because timing-offset estimation is equivalent to estimating a fractional delay on this grid²⁰.

When the receiver also performs any segmentation of long captures into blocks (e.g., per-OSNR traces or per-burst records), the same “time-first” principle applies: segmentation boundaries should be defined in seconds (or in symbol counts after

¹⁸ S. Jang, J. Lee, Y. Choi, D. Kim, G. Kim, “Recent Advances in Ultrahigh-Speed Wireline Receivers With ADC-DSP-Based Equalizers”, IEEE Open Journal of the Solid-State Circuits Society, vol. 4, pp. 290–304, 2024.

¹⁹ *Ibidem*.

²⁰ A. Nasir, S. Durrani, H. Mehrpouyan, S. Blostein, R. Kennedy, “Timing and carrier synchronization in wireless communication systems: a survey and classification of research in the last 5 years”, EURASIP Journal on Wireless Communications and Networking, 180, 2016.

an explicit R_s mapping) so that subsequent symbol alignment remains invariant to the chosen working rate or any later rate conversions. This is especially important when comparing multiple acquisitions where the oscilloscope clock and trigger alignment may vary slightly across runs²¹.

7.2.2 Timing recovery & symbol synchronization (high-level + key equations)

After resampling, timing recovery aims to estimate the sampling phase (fractional symbol timing) that yields symbol-synchronous samples. A standard discrete-time viewpoint models the working-rate sequence as samples of an underlying continuous-time received waveform $r(t)$ taken at times

$$t_m = mT_s + \tau,$$

where $T_s = 1/F_{s,\text{work}}$ and τ is an unknown timing offset to be tracked (possibly time-varying). Timing recovery is commonly implemented as a feedback loop that updates τ using a timing error detector (TED) operating on interpolated samples²².

For fractionally spaced operation at (2) Sa/sym, a widely used non-data-aided TED is the Gardner detector. Denote the symbol-spaced “center” sample by x_k and the half-symbol “early/late” samples by $x_{k\pm 1/2}$ (all obtained by interpolation from the working-rate sequence). The (real-valued) Gardner timing error can be written as

$$e_k^G = (x_{k+1/2} - x_{k-1/2})x_k,$$

which is then filtered and used to update the timing estimate, e.g.,

$$\tau_{k+1} = \tau_k + \mu e_k,$$

with loop gain $\mu > 0$ (and, in practice, a loop filter between e_k and the update). The

²¹ C. Truong, L. Oudre, N. Vayatis, “*Selective review of offline change point detection methods*”, Signal Processing, vol. 167, Art. 107299, 2020.

²² *Ibidem*. A. Nasir *et al.*

key advantage in PAM4-style receivers is that this TED naturally exploits 2 Sa/sym structure and does not require hard symbol decisions during acquisition²³.

A common decision-directed alternative is the Mueller–Müller (M&M) TED, which uses detected symbols $\hat{a}_k$ (from a slicer or preliminary equalizer). One standard real-valued form is

$$e_k^{\text{MM}} = x_k \hat{a}_{k-1} - x_{k-1} \hat{a}_k,$$

again combined with a loop filter and timing update. In practice, IM/DD PAM4 systems often use a staged approach: a coarse synchronization (e.g., correlation to a known training header or PRBS alignment) initializes the frame and approximate sampling phase, and then a TED-based loop refines τ while the equalizer converges²⁴.

Finally, it is essential to note that receiver filtering and bandwidth constraints directly affect TED behavior: excessive pre-filtering can reduce timing-gradient information, whereas insufficient filtering increases jitter in e_k through additive noise. Therefore, timing recovery and front-end filtering are not independent design choices in a bandwidth-limited PAM4 receiver chain²⁵.

7.2.3 DC removal/normalization/gain control (AGC-like steps)

IM/DD receivers generate a real, nonnegative electrical waveform after square-law detection and TIA/AFE amplification; practical captures often include a significant DC component (due to average optical power and receiver bias) and slow baseline drift (caused by thermal effects, bias variations, or low-frequency noise processes). If not addressed, DC and baseline terms can overshadow the record's variance and impair both timing recovery and equalizer adaptation by amplifying low-frequency energy that does not provide useful information for symbol decisions. Therefore, a common front-end step involves subtracting a baseline estimate, such as a blockwise mean or a slow trend estimate, resulting in

²³ *Ibidem.*

²⁴ *Ibidem.*

²⁵ *Ibidem.*

$$\tilde{r}[n] = r[n] - \hat{b}[n],$$

where $\hat{b}[n]$ may be constant per block (mean removal) or slowly time-varying (detrending/low-cut filtering)²⁶.

After DC removal, amplitude normalization ensures a consistent numerical range across OSNR points and acquisitions. A typical practice is to scale by a robust power estimate (e.g., RMS of the zero-mean trace over a training portion) so that $\text{Var}(\tilde{r})$ is approximately constant across records:

$$r_{\text{norm}}[n] = \frac{\tilde{r}[n]}{\sqrt{\mathbb{E}\{\tilde{r}^2[n]\}}}.$$

This stabilization is practically important because adaptive algorithms (FFE adaptation, timing loops, and any subsequent nonlinear processing) are step-size-sensitive; normalization reduces the need to retune hyperparameters for each measurement condition²⁷.

In hardware-focused receiver discussions (wireline and related high-speed ADC-DSP front ends), gain control is also described as range management for quantization: the effective gain and scaling should position the signal within the usable ADC range, preventing clipping while preserving sufficient resolution. Even in offline optical experiments (where the scope sets the ADC gain), an equivalent “digital AGC” perspective applies: digital scaling is selected to avoid numerical saturation and keep adaptation stable, mirroring front-end AGC considerations in other receiver types²⁸.

²⁶ C. Caballero-Gaudes, R. Reynolds, “Methods for cleaning the BOLD fMRI signal”, *NeuroImage*, vol. 154, pp. 128–149, 2017.

²⁷ S. Jang, J. Lee, Y. Choi, D. Kim, G. Kim, “Recent Advances in Ultrahigh-Speed Wireline Receivers With ADC-DSP-Based Equalizers”, *IEEE Open Journal of the Solid-State Circuits Society*, vol. 4, pp. 290–304, 2024.

²⁸ A. Kumari, D. Bhatt, “Advanced System Analysis and Survey on the GPS Receiver Front End”, *IEEE Access*, vol. 10, pp. 24611–24626, 2022.

7.2.4 Matched filtering vs practical receiver filtering choices in IM-DD

In classical AWGN theory, the matched filter to a known transmit pulse maximizes SNR at the sampling instant and provides sufficient statistics for symbol detection. However, in practical IM/DD PAM4 links, the effective received pulse is shaped by the combined effects of transmitter pulse shaping, modulator and electrical bandwidth limits, photodiode and TIA response, and any analog anti-alias filtering. Often, the receiver does not have an accurately known, time-invariant impulse response. As a result, many experimental PAM4 receiver chains do not implement a strict matched filter; instead, they use practical low-pass filtering, either analog or digital, to suppress out-of-band noise and prevent aliasing at the chosen operating rate. This leaves residual ISI to be managed by fractionally spaced equalization²⁹.

This “practical filtering first, equalization second” approach reflects broader front-end design tradeoffs observed in oversampled ADC-DSP receivers: analog filtering is constrained by manufacturability and power consumption. In contrast, digital filtering must balance noise reduction against increased memory and intersymbol interference (ISI), which raises equalizer complexity. In bandwidth-limited IM/DD systems, aggressive low-pass filtering can enhance noise robustness but also deepen spectral notches and increase channel memory, shifting more responsibility onto the equalizer and detector³⁰.

From a receiver architecture perspective, the matched-filter concept remains a useful reference. If an effective pulse response (including AFE) can be estimated, one can design a digital filter that approximates a matched filter (or a whitening matched filter) to enhance detector performance. However, in many practical front ends, the filtering stage is best regarded as part of the overall synchronization and equalization co-design: filtering choices affect timing-loop behavior and the statistics encountered by adaptation, so they are selected to ensure the overall chain

²⁹ S. Jang, J. Lee, Y. Choi, D. Kim, G. Kim, “Recent Advances in Ultrahigh-Speed Wireline Receivers With ADC-DSP-Based Equalizers”, IEEE Open Journal of the Solid-State Circuits Society, vol. 4, pp. 290–304, 2024.

³⁰ S. Jang, J. Lee, Y. Choi, D. Kim, G. Kim, “Recent Advances in Ultrahigh-Speed Wireline Receivers With ADC-DSP-Based Equalizers”, IEEE Open Journal of the Solid-State Circuits Society, vol. 4, pp. 290–304, 2024.

remains stable and implementable rather than to achieve optimal isolation³¹ in theory.

7.3 The Equalizer's Role in IM-DD DSP

Equalization is the key DSP function in short-reach IM-DD PAM4 because the end-to-end electrical-optical-electrical path is inherently bandwidth-limited and dispersive. Additionally, square-law direct detection converts phase-related impairments into intensity distortions that manifest as intersymbol interference (ISI) and noise amplification after inversion. The equalizer thus acts as the receiver's tool for approximately reversing a memory channel within practical limits, such as limited oversampling, tap length, complexity, and imperfect impairment knowledge, while balancing residual distortion, noise boost, and decision errors. A thorough approach views this as an inverse problem: the receiver samples a distorted, noisy waveform. It aims to reconstruct the transmitted symbols (hard decisions) or their posterior information (soft decisions) with acceptable robustness and complexity³².

7.3.1 Why equalization is unavoidable (ISI sources across driver-MZM-PD/TIA + bandwidth limits)

In datacenter-style IM-DD interconnects, the “channel” relevant to DSP is the cascade of transmitter and receiver electrical bandwidth (DAC/driver, modulator electrical response, photodiode/TIA and oscilloscope/ADC front-end), optionally combined with residual fiber chromatic dispersion (CD) even at short distances (connectors, patch cords, laboratory spans) and with any intentional bandwidth shaping. System-level analyses emphasize that, at high baud rates, this cascade produces a severely band-limited effective impulse response with non-negligible memory, so that symbol-rate sampling without equalization yields closed eyes and

³¹ A. Nasir, S. Durrani, H. Mehrpouyan, S. Blostein, R. Kennedy, “Timing and carrier synchronization in wireless communication systems: a survey and classification of research in the last 5 years”, EURASIP Journal on Wireless Communications and Networking, 180, 2016.

³² D. Che, X. Chen, “Modulation Format and Digital Signal Processing for IM-DD Optics at Post-200G Era”, Journal of Lightwave Technology, vol. 42, pp. 588-605, 2024.

large penalties; consequently, practical PAM4 links either allocate DSP for post-equalization or shift part of the inversion to the transmitter via pre-equalization to maintain feasibility at scale³³.

When dispersion occurs in double-sideband IM-DD links, square-law detection can convert dispersion-induced phase rotation into frequency-dependent intensity fading, often referred to as power fading, which further exacerbates ISI and may create deep spectral notches. Experimental research on dispersion-uncompensated transmission shows that, even with careful pulse shaping, residual CD and bandwidth limits require receiver-side equalization. Performance recovery depends on compensating for the resulting memory and distortion, rather than treating the received samples as memoryless PAM observations³⁴.

A particularly practical manifestation of “unavoidability” is that aggressive linear equalization of a bandwidth-limited channel tends to amplify colored noise in spectral dips (noise enhancement), so the equalizer is not only an ISI remover but also a design trade-off point between distortion suppression and noise boosting. This motivates receiver-side structures that explicitly address the post-equalization noise coloration (e.g., noise-whitening-aware processing) rather than applying a naïve inverse filter³⁵.

7.3.2 Equalizer as an inverse problem (linear vs nonlinear channel view)

A compact discrete-time representation suitable for PAM4 receiver DSP is to treat the post-detection, sampled waveform as the output of a finite-memory operator driven by the transmitted symbol sequence. In its linearized form (useful as a baseline and for many practical designs), one writes a fractionally-sampled model

³³ X. Zhou, R. Urata, H. Liu, “Beyond 1 Tb/s Intra-Data Center Interconnect Technology: IM-DD OR Coherent?”, *Journal of Lightwave Technology*, vol. 38, pp. 475-484, 2020.

³⁴ X. Wu, J. Zhang, A. Lau, C. Lu, “C-band 120-Gb/s PAM-4 transmissions over a 100-km dispersion-uncompensated SSMF using joint combined pulse shaping and low-complexity nonlinear equalization”, *Optics Letters*, vol. 47, no. 19, pp. 5144-5147, 2022.

³⁵ Q. Wu, Z. Xu, Y. Zhu, H. Ji, Y. Yang, G. Qiao, L. Liu, S. Wang, J. Liang, J. Wei, Z. He, Q. Zhuge, W. Hu, “Gradient-descent noise whitening techniques for short-reach IM-DD optical interconnects with severe bandwidth limitation”, *Optics Express*, vol. 32, no. 2, pp. 1715-1727, 2024.

$$y[m] = \sum_{k=0}^{L_h-1} h[k] a[m-k] + w[m],$$

where $a[\cdot]$ are PAM4 symbols (or oversampled pulse-shaped samples derived from them), $h[\cdot]$ is the effective discrete-time impulse response of the complete path, and $w[m]$ aggregates thermal/quantization noise and other approximately additive terms. In this view, linear equalization seeks an approximate inverse of $h[\cdot]$ (e.g., ZF/MMSE-FFE) to recover $a[\cdot]$. Still, it must manage the inevitable noise enhancement when $H(e^{j\omega})$ has small magnitude over portions of the band³⁶.

However, IM-DD PAM4 links frequently depart from strict linearity: square-law detection, modulator transfer characteristics, and bandwidth-limited devices can yield nonlinear ISI and signal-dependent distortions that are not captured by a single linear convolution model. One widely used system-identification perspective is therefore to represent the receiver observation as

$$y[m] = \mathcal{F}(a[m], a[m-1], \dots, a[m-L+1]) + v[m],$$

where $\mathcal{F}(\cdot)$ is a nonlinear finite-memory mapping (often approximated by Volterra-like expansions or simplified nonlinear bases) and $v[m]$ is residual noise. Low-complexity nonlinear equalizers based on selected nonlinear terms (e.g., absolute-term-based structures) explicitly adopt this “nonlinear inverse problem” viewpoint by choosing a restricted nonlinear basis that targets dominant IM-DD nonlinearities while controlling the number of coefficients³⁷.

A complementary inverse-problem formulation is sequence detection: rather than producing per-symbol decisions after linear equalization, one models the channel memory explicitly and performs reduced-state maximum-likelihood

³⁶ D. Che, X. Chen, “Modulation Format and Digital Signal Processing for IM-DD Optics at Post-200G Era”, *Journal of Lightwave Technology*, vol. 42, pp. 588-605, 2024.

³⁷ J. Zhang, Z. Lin, X. Wu, J. Liu, A. Lau, C. Guo, C. Lu, S. Yu, “Low-complexity sparse absolute-term based nonlinear equalizer for C-band IM/DD systems”, *Optics Express*, vol. 29, no. 14, pp. 21891-21901, 2021.

sequence estimation (MLSE) on the PAM sequence, trading state complexity against performance. This approach is particularly natural when the channel can be well approximated as a short-memory ISI process and when reliable branch metrics can be derived after front-end conditioning³⁸.

For severe ISI regimes, probabilistic trellis-based inference can be pushed further toward BCJR-style processing, in which assistance mechanisms (e.g., least-squares-aided metric formation) are used to stabilize operation under harsh bandwidth limitations. This again fits the inverse-problem framing: equalization is not merely “filtering,” but an inference procedure under a structured memory model³⁹.

7.3.3 Placement options (pre-eq/post-eq; symbol-rate vs fractionally-spaced)

Equalization can be placed (i) at the transmitter as digital pre-equalization, (ii) at the receiver as post-equalization, or (iii) jointly, splitting compensation between both ends. In IM-DD links, transmitter pre-equalization is often motivated by the desire to partially invert the transmitter-side bandwidth limitation (DAC/driver/modulator path) before optical modulation, thereby reducing the burden on receiver DSP and mitigating downstream noise enhancement. Comparative studies of joint linear and nonlinear pre-equalization schemes in IM/DD access/interconnect contexts show that moving part of the compensation upstream can improve performance, but also introduces sensitivity to transmitter operating point and requires careful management of clipping/quantization constraints⁴⁰.

Joint pre- and post-equalization is attractive when neither end alone can implement a stable, low-noise inversion over the full effective bandwidth. In this

³⁸ Y. Yu, Y. Che, T. Bo, D. Kim, H. Kim, “*Reduced-state MLSE for an IM/DD system using PAM modulation*”, Optics Express, vol. 28, no. 26, pp. 38505-38515, 2020.

³⁹ S. Liu, J. Ming, J. Song, Y. Mu, H. He, R. Wang, R. Gao, Z. Li, X. Xin, Z. Dong, “*Least-squares-assisted BCJR algorithm for severe ISI mitigation in bandwidth-limited IM/DD optical interconnection*”, Optics Express, vol. 33, no. 12, pp. 24545-24552, 2025.

⁴⁰ G. Li, Z. Li, Y. Ha, F. Hu, J. Zhang, N. Chi, “*Performance Assessments of Joint Linear and Nonlinear Pre-Equalization Schemes in Next Generation IM/DD PON*”, Journal of Lightwave Technology, vol. 40, pp. 5478-5489, 2022.

partitioned view, a short pre-equalizer can shape the launched waveform to reduce the most harmful ISI components, while a receiver post-equalizer removes residual memory and improves numerical conditioning (and thus reduces noise enhancement compared to a purely receiver-side ZF-like inversion). This placement logic is explicitly treated in joint pre/post frameworks for multilevel signaling, emphasizing the engineering trade-off between where complexity is paid and how robust the overall inverse is⁴¹.

Within receiver post-equalization, a key architectural choice is between symbol-rate and fractionally spaced processing. Symbol-rate equalizers operate on one sample per symbol and can be efficient when timing is tightly controlled and the effective channel is well-conditioned at the sampling phase. Fractionally-spaced equalizers (commonly at 2 Sa/sym in offline IM-DD work) operate on oversampled sequences, improving tolerance to sampling-phase error and allowing the equalizer to jointly address timing-phase sensitivity and ISI in a unified linear structure, at the cost of higher tap count and computation. Surveys of IM-DD DSP emphasize fractionally spaced processing as a practical baseline under bandwidth limitations and implementation constraints⁴².

Decision-feedback equalization (DFE) is often introduced as a placement refinement: by feeding back past decisions, DFE can mitigate post-cursor ISI with reduced noise enhancement relative to long feed-forward inverses, but it introduces error propagation and, in bandwidth-limited IM-DD links, may generate burst-error behavior that complicates soft-decision FEC interfaces. This motivates hybrid detection strategies that explicitly address DFE burst errors while maintaining low complexity⁴³.

⁴¹ J. Singh, A. Sandmann, A. Ahrens, S. Lochmann, “Joint pre- and post-equalization using optical multi-level signaling”, *The Journal of Engineering*, Vol. 2022, Issue4, pp. 442-446, 2022.

⁴² D. Che, X. Chen, “Modulation Format and Digital Signal Processing for IM-DD Optics at Post-200G Era”, *Journal of Lightwave Technology*, vol. 42, pp. 588-605, 2024.

⁴³ K. Wu, G. Liga, J. Riani, A. Alvarado, “Low-Complexity Soft-Decision Detection for Combating DFE Burst Errors in IM/DD Links”, *Journal of Lightwave Technology*, vol. 42, pp. 1395-1408, 2024.

7.3.4 Performance metrics used to judge equalization (BER, EVM, eye metrics, robustness)

In IM-DD PAM4, the most direct equalization performance metric is BER, typically reported against hard-decision FEC thresholds (and, where available, against soft-decision relevant targets). Equalization quality is also commonly assessed using receiver sensitivity or the required OSNR/received optical power at a target BER, since equalization alters both residual ISI and the effective noise statistics after processing. System-level IM-DD discussions use these metrics to quantify whether a given equalizer architecture meaningfully shifts the power/OSNR operating point for a target baud rate and reach⁴⁴.

EVM is frequently used as a modulation-quality proxy because it captures the RMS deviation of equalized samples from ideal constellation levels and is sensitive to residual linear distortion, nonlinear warping, and noise enhancement. While BER is the ultimate decision metric, EVM provides a diagnostic link between equalizer settings (tap length, step size, regularization) and the quality of the equalized symbol clusters, especially when comparing linear and nonlinear compensation under the same acquisition conditions⁴⁵.

Eye-diagram metrics (eye height, eye width, crossing symmetry, and timing-jitter sensitivity) remain valuable, particularly for interpreting how an equalizer reshapes the waveform and for distinguishing distortion-dominated from noise-dominated regimes. In practice, robustness is judged by how stable BER/EVM/eye metrics remain under bandwidth variation, modest CD mismatch, and drift in component responses, precisely the regimes where design choices such as noise-whitening-aware processing or burst-error-mitigating detection materially affect the reliability of the DSP chain⁴⁶.

⁴⁴ X. Zhou, R. Urata, H. Liu, “Beyond 1 Tb/s Intra-Data Center Interconnect Technology: IM-DD OR Coherent?”, *Journal of Lightwave Technology*, vol. 38, pp. 475-484, 2020.

⁴⁵ D. Che, X. Chen, “Modulation Format and Digital Signal Processing for IM-DD Optics at Post-200G Era”, *Journal of Lightwave Technology*, vol. 42, pp. 588-605, 2024.

⁴⁶ Q. Wu, Z. Xu, Y. Zhu, H. Ji, Y. Yang, G. Qiao, L. Liu, S. Wang, J. Liang, J. Wei, Z. He, Q. Zhuge, W. Hu, “Gradient-descent noise whitening techniques for short-reach IM-DD optical interconnects with severe bandwidth limitation”, *Optics Express*, vol. 32, no. 2, pp. 1715-1727, 2024.

7.4 Classical Equalizers for IM-DD PAM4

Classical (non-ML) equalizers remain the baseline in IM/DD PAM4 receivers because they directly target the dominant impairment in bandwidth-limited links: intersymbol interference (ISI) created by the concatenated frequency responses of the DAC/driver, modulator, photodiode/TIA, and practical receive filtering (and, depending on reach, dispersion-related power fading). In system-level DSP, these equalizers are typically formulated as (i) linear feed-forward filtering (FFE), optionally at fractional spacing, (ii) decision-feedback equalization (DFE) to cancel post-cursor ISI without fully inverting spectral dips, (iii) adaptive identification of the equalizer coefficients (training or decision-directed), (iv) expanded nonlinear models (Volterra / memory-polynomial families) when square-law and device nonlinearities are non-negligible, and (v) sequence detection (MLSE/Viterbi) as a near-optimum reference for finite-memory ISI channels⁴⁷.

7.4.1 Linear equalizers: FFE (T-spaced / fractionally-spaced), MMSE/ZF views

Channel abstraction and FFE structure.

A common starting point is a discrete-time ISI model

$$r[n] = \sum_{k=0}^{L_h-1} h[k] a[n-k] + w[n],$$

where $a[n]$ are PAM4 symbols, $h[k]$ captures the effective bandwidth-limited response (often including CD-induced shaping in IM/DD), and $w[n]$ is an aggregate noise term. Linear equalization applies an FIR filter $g[k]$ to produce

$$y[n] = \sum_{k=0}^{L_g-1} g[k] r[n-k],$$

with coefficients selected to reduce ISI and enable low-complexity symbol

⁴⁷ J. G. Proakis and M. Salehi, “*Digital Communications*,” 5th ed., McGraw-Hill, 2007.

decisions. This is the core linear equalization view used across short-reach IM/DD interconnect discussions⁴⁸.

Symbol-spaced vs fractionally-spaced FFE (FSE).

A symbol-spaced FFE samples at (approximately) one sample per symbol and uses one-tap-per-symbol delays. In bandwidth-limited IM/DD PAM4, however, receivers often operate at 2 samples/symbol (T/2 spacing) (via offline oscilloscope capture or ADC oversampling), enabling a fractionally spaced equalizer (FSE) that is more robust to timing phase errors and better approximates inversion when the effective channel exhibits significant spectral distortion. The use of a T/2-spaced post-FFE is explicitly reported as the residual-ISI mitigation stage in pre-electronic-dispersion-compensated IM/DD links, where much of the dispersion is handled before detection and a modest-length FSE removes remaining ISI⁴⁹.

Hybrid spacing as a complexity lever.

At very high rates, implementing a full T/2 FSE can be costly; a pragmatic compromise is a hybrid FFE that combines symbol-spaced and fractionally spaced taps to preserve most of the benefit of T/2 equalization at reduced arithmetic cost. An experimental demonstration of such a low-complexity hybrid FFE for strictly band-limited IM/DD systems reports near-equivalent BER performance to a conventional T/2 FFE with substantially lower implementation complexity⁵⁰.

ZF and MMSE interpretations (linear inverse problem).

⁴⁸ X. Pang, W. Hu, G. Jacobsen, S. Popov, J. Chen, O. Ozoliņš, R. Lin, L. Zhang, A. Udalcovs, L. Xue, R. Schatz, U. Westergren, and S. Xiao, “200 Gbps/Lane IM/DD Technologies for Short Reach Optical Interconnects,” *Journal of Lightwave Technology*, 38, 492–503, 2020.

⁴⁹ X. Wu, A. Karar, K. Zhong, A. Lau, and C. Lu, “Experimental demonstration of pre-electronic dispersion compensation in IM/DD systems using an iterative algorithm,” *Optics Express*, 29(16), 24735–24749, 2021

⁵⁰ J. Huo, M. Sun, X. Zhou, H. Wei, J. Yuan, H. Ning, and K. Long, “Experimental demonstration of low complexity hybrid FFE algorithm for strictly band-limited IM/DD system,” *Optics Communications*, Vol. 505, 127485 2022.

Linear equalization can be framed as an inverse problem on the ISI channel. In the frequency domain, a zero-forcing (ZF) equalizer aims to invert the channel response, ideally $G_{\text{ZF}}(e^{j\omega}) \approx 1/H(e^{j\omega})$, which removes ISI but can strongly enhance noise near spectral dips. In contrast, a minimum mean-square error (MMSE) equalizer chooses $g[k]$ to minimize $\mathbb{E}\{|a[n] - y[n]|^2\}$, balancing residual ISI against noise enhancement; in WSS settings, this yields the classical Wiener form

$$G_{\text{MMSE}}(e^{j\omega}) = \frac{H^*(e^{j\omega})S_a(e^{j\omega})}{|H(e^{j\omega})|^2 S_a(e^{j\omega}) + S_w(e^{j\omega})}.$$

These ZF/MMSE viewpoints are standard in digital communication equalizer theory and remain the appropriate conceptual lens for mapping IM/DD bandwidth limits into a discrete-time ISI model⁵¹.

Where linear FFE appears in practical IM/DD PAM4 implementations.

Linear equalization is not only a receiver-DSP block; it also appears as transmitter pre-emphasis/pre-FFE in CMOS PAM4 transmitters, where fractional-spaced FIR structures are integrated with the driver to shape the emitted waveform and ease receiver requirements. A PAM4 transmitter in 65-nm CMOS explicitly reports a fractional-spaced FFE architecture to support 50-112 Gb/s operation, illustrating how “equalization” can be partly pushed into the front-end hardware⁵². A later 28-nm CMOS PAM4 transmitter further reports a reconfigurable three-tap $T/2$ – T variable fractional-spaced FFE, reinforcing that fractional spacing is a practical design dimension for bandwidth-limited PAM4 links⁵³.

⁵¹ J. G. Proakis and M. Salehi, “*Digital Communications*,” 5th ed., McGraw-Hill, 2007.

⁵² X. Zheng, H. Ding, F. Zhao, D. Wu, L. Zhou, J. Wu, F. Lv, J. Wang, and X. Liu, “A 50–112-Gb/s PAM-4 Transmitter With a Fractional-Spaced FFE in 65-nm CMOS,” *IEEE Journal of Solid-State Circuits*, 55, 1864–1876, 2020.

⁵³ K. Wang, R. Tang, S. Xiang, Y. He, Y. He, and X. Gui, “A 0.92-pJ/b 112-Gb/s PAM-4 Transmitter With Bandwidth and Linearity Enhanced Quasi-Voltage-Mode Driver and Reconfigurable Three-Tap $T/2$ – T Variable Fractional-Spaced FFE in 28-nm CMOS,” *IEEE Transactions on Circuits and Systems I: Regular Papers*, 72, 2664–2675, 2025.

Finally, linear equalization can be partitioned across optical/electrical domains. A joint optical–electrical feedforward equalization (OE-FFE) approach analyzes and demonstrates low-complexity equalization where part of the compensation is realized optically and the remainder electrically/digitally, which is especially relevant when bandwidth/dispersion constraints make pure electronic inversion inefficient⁵⁴.

7.4.2 Decision-feedback equalizer (DFE): benefits and error propagation

Structure and benefit.

A DFE augments a feed-forward filter with a feedback filter driven by past detected symbols:

$$y[n] = \sum_{k=0}^{L_f-1} f[k] r[n-k] - \sum_{k=1}^{L_b} b[k] \hat{a}[n-k].$$

The feedback path cancels *post-cursor* ISI using decisions rather than directly inverting the channel spectrum, so it can mitigate severe post-cursor ISI with less noise enhancement than a pure ZF-like FFE, which is an attractive feature in bandwidth-limited PAM4 links⁵⁵.

Error propagation and burst errors.

The principal drawback is error propagation: a wrong decision $\hat{a}[n-k]$ is fed back and can corrupt subsequent decisions, producing burst errors. This mechanism is not merely theoretical; it directly motivates receiver designs that explicitly address DFE-induced burst errors and soft-decision reliability in IM/DD links,

⁵⁴ P. Zhu, Y. Yoshida, A. Kanno, and K. Kitayama, “*Analysis and Demonstration of Low-Complexity Joint Optical-Electrical Feedforward Equalization (OE-FFE) For Dispersion-Limited High-Speed IM/DD Transmission*,” *Journal of Lightwave Technology*, 41, 477–488, 2023.

⁵⁵ J. G. Proakis and M. Salehi, “*Digital Communications*,” 5th ed., McGraw-Hill, 2007.

where bursts can be particularly damaging for modern FEC interaction and performance prediction⁵⁶.

Reliability-aware and structurally modified DFEs (IM/DD examples).

To suppress error propagation while keeping DFE's ISI-cancellation benefit, several IM/DD-oriented modifications have been reported. An improved weighted DFE for 120-Gb/s PAM4 over 50-km SSMF scales feedback using reliability-related weighting to reduce propagation and improve performance⁵⁷. A different approach maps the PAM4 signal into a complex-valued processing domain, proposing a complex-valued DFE with adaptive manipulations in time and amplitude domains for optical IM/DD signals⁵⁸. For VCSEL-based short-reach IM/DD, a transition-based DFE has been proposed to enhance transmission performance under strong bandwidth constraints, reflecting a practical strategy to reduce burst behavior in realistic short-reach channels⁵⁹.

Hybridization with sequence detection.

Another pragmatic line is to combine DFE with selective sequence processing only when needed. An MLSE-assisted DFE is reported for suppressing error propagation in high-speed IM/DD transmission systems, aiming to retain a DFE-

⁵⁶ K. Wu, G. Liga, J. Riani, and A. Alvarado, "Low-Complexity Soft-Decision Detection for Combating DFE Burst Errors in IM/DD Links," *Journal of Lightwave Technology*, 42, 1395–1408, 2023.

⁵⁷ J. Zhang, X. Wu, L. Sun, J. Liu, A. Lau, C. Guo, S. Yu, and C. Lu, "C-band 120-Gb/s PAM-4 transmission over 50-km SSMF with improved weighted decision-feedback equalizer," *Optics Express*, Vol. 29, Issue 25, pp. 41622–41633, 2021.

⁵⁸ L. Sun, J. Xiao, Y. Cai, G. Shen, N. Liu, and C. Lu, "Complex-valued decision feedback equalizer for optical IMDD signals with adaptive manipulations in time and amplitude domains," *Optics Letters*, 47(17), 4391–4394, 2022.

⁵⁹ B. Widhianto, H. Tsui, and J. Chen, "Transition-Based Decision Feedback Equalization to Enhance Transmission in VCSEL-Based Short-Reach IM/DD System," *IEEE Photonics Journal*, 17, 7201410, 2025.

like baseline cost while invoking higher-complexity detection mechanisms in the regimes where bursts are most harmful⁶⁰.

7.4.3 Adaptive algorithms (LMS/RLS; decision-directed variants; practical notes)

Training vs decision-directed adaptation.

In practice, equalizer coefficients must be learned from data. Classical receivers commonly use a training mode (known pilot symbols) to initialize coefficients and, when needed, a decision-directed (DD) mode to continue tracking without continuous pilots. The standard adaptive equalization formulation minimizes a sample MSE using an error signal $e[n] = d[n] - y[n]$, where $d[n]$ denotes a known training symbol or a detected symbol in DD mode. These operating modes and their trade-offs (convergence, steady-state misadjustment, and error propagation under DD) are canonical results in digital-communication equalizer theory⁶¹.

LMS family (low complexity).

For an FFE or the feed-forward portion of a DFE, the complex-baseband LMS update is

$$g_{n+1} = g_n + \mu e[n] r_n^*,$$

where r_n is the tapped delay-line vector and μ is the step size. LMS is computationally light and robust, which explains its pervasive use as a baseline adaptation mechanism even when final system designs incorporate additional refinements⁶².

RLS family (fast convergence).

⁶⁰ J. Zhou, J. Zhang, X. Zhao, R. Wang, J. Li, S. Hu, Q. Zhang, Q. Yang, B. Xu, and K. Qiu, “*MLSE-assisted decision feedback equalizer for error-propagation suppression in high-speed IM-DD transmission systems*,” Chinese Optics Letters, 2024.

⁶¹ J. G. Proakis and M. Salehi, “*Digital Communications*,” 5th ed., McGraw-Hill, 2007.

⁶² *Ibidem*.

RLS updates coefficients using an adaptive estimate of the inverse correlation matrix to achieve much faster convergence (at higher per-sample cost). The classical RLS recursion is often summarized as

$$g_n = g_{n-1} + K_n e[n], \quad K_n = \frac{P_{n-1} r_n}{\lambda + r_n^H P_{n-1} r_n},$$

with P_n updated accordingly and λ a forgetting factor. The LMS–RLS trade-off (complexity vs tracking speed) is again a standard foundation for discussing practical equalizer adaptation⁶³.

PAM4-specific practicalities: tap and threshold adaptation.

Multilevel PAM receivers often require both tap adaptation and decision-threshold adaptation because the effective scaling and offset of PAM levels can drift due to front-end gain, clipping, and bandwidth-dependent distortion. A PAM4 transceiver design that explicitly combines threshold-adaptive and tap-adaptive mechanisms demonstrates this practical receiver reality: the equalizer “algorithm” involves both filter-coefficient updates and decision-level control⁶⁴.

Implementation perspective from high-speed serial links.

Although optical IM/DD imposes its own constraints, high-speed serial-link literature offers valuable implementation insights for DFEs: timing closure, feedback latency, and calibration complexity often take precedence. A comprehensive review of decision-feedback equalization techniques for multi-gigabit serial data links highlights these engineering challenges and corresponding mitigation strategies that apply to high-speed PAM4 optical receivers⁶⁵. A more

⁶³ *Ibidem*.

⁶⁴ X. Liu, Z. Li, H. Wen, M. Miao, Y. Wang, and Z. Wang, “A PAM4 transceiver design scheme with threshold adaptive and tap adaptive,” *EURASIP Journal on Advances in Signal Processing*, 2023, 70, 2023.

⁶⁵ F. Yuan, A. Al-Taei, A. Ye, and S. Sadr, “Design techniques for decision feedback equalisation of multi-giga-bit-per-second serial data links: a state-of-the-art review,” *IET Circuits, Devices & Systems*, 8, 118–130, 2014.

hardware-calibration-oriented example is an embedded DDR5 DFE calibration architecture based on channel stimulus inversion, highlighting how practical designs may rely on structured calibration to avoid unstable online adaptation regimes⁶⁶.

7.4.4 Nonlinear/expanded models (e.g., Volterra-type equalization, memory-polynomial views)

Why linear models can be insufficient in IM/DD.

In IM/DD, square-law detection and device nonlinearities (driver/modulator transfer, photodiode/TIA non-idealities) can cause distortion that is not accurately modeled by a purely linear ISI model, particularly when combined with dispersion effects. Classical nonlinear equalization addresses this by expanding the model class beyond linear FIRs while maintaining structure and interpretability.

Volterra and simplified polynomial structures.

A discrete-time Volterra equalizer expresses $y[n]$ as a sum of polynomial functionals of delayed samples (including cross-terms). Full Volterra models become computationally expensive as order and memory requirements increase, motivating structured simplifications. A low-complexity, sparse, absolute-term-based nonlinear equalizer is reported for C-band IM/DD systems, explicitly targeting nonlinear distortion while reducing arithmetic complexity relative to richer Volterra forms⁶⁷. A follow-on comparison study analyzes sparse and weight-sharing nonlinear equalizers for 100-Gbit/s DSB PAM4 over 60-km SSMF, reflecting the core classical engineering goal: retain most of the nonlinear compensation benefit while compressing multiplications and memory⁶⁸.

⁶⁶ M. Cooke and N. Nicolici, “An Embedded Architecture for DDR5 DFE Calibration Based on Channel Stimulus Inversion,” *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 33, 793–806, 2025.

⁶⁷ X. Liu, Y. Liu, C. Hu, and J. Zhao, “Experimental demonstration of low-complexity trigonometric-memory-polynomial decision-feedback equalizer for nonlinear compensation in IM/DD optical systems,” *Optics Express*, 32(3), 3329–3341, 2024.

⁶⁸ X. Liu, C. Hu, Y. Cai, and J. Zhao, “Exponential-Memory-Polynomial-Based Nonlinear Equalizer for C-Band IM/DD Systems,” *IEEE Photonics Technology Letters*, 36, 1261–1264, 2024.

Memory-polynomial families: trigonometric/exponential variants.

A common nonlinear-equalization compromise is the memory polynomial form, which limits nonlinearities to per-tap magnitude-dependent terms (reducing cross-term explosion). Recent IM/DD-focused research demonstrates a low-complexity trigonometric-memory-polynomial DFE experimentally, positioning it as a structured nonlinear equalizer with advantageous complexity–performance trade-offs⁶⁹. Related work proposes an exponential-memory-polynomial-based nonlinear equalizer for C-band IM/DD systems, again emphasizing implementable nonlinear modeling rather than unconstrained black-box compensation⁷⁰.

Physics-informed structure.

Beyond generic polynomials, physics-informed constraints can further compress the model. A computationally efficient physics-informed difference-symmetric nonlinear equalizer is proposed for C-band DML-DD links, explicitly embedding structural symmetry to reduce computation while maintaining nonlinear compensation capability⁷¹.

Decision-directed nonlinear equalization in bandwidth-limited IMDD.

Nonlinear equalizers can also be trained in a decision-directed way. A study on IMDD bandwidth-limited PAM8 transmission shows that performance can be improved with a tap-coefficient decision-directed Volterra equalizer, demonstrating

⁶⁹ X. Liu, C. Hu, Y. Cai, and J. Zhao, “Exponential-Memory-Polynomial-Based Nonlinear Equalizer for C-Band IM/DD Systems,” *IEEE Photonics Technology Letters*, 36, 1261–1264, 2024.

⁷⁰ Y. Yu, Y. Che, T. Bo, D. Kim, and H. Kim, “Reduced-state MLSE for an IM/DD system using PAM modulation,” *Optics Express*, 28(26), 38505–38515, 2020.

⁷¹ Z. Chen, J. Nie, S. Zhang, Q. Yang, X. Dai, L. Deng, M. Cheng, and D. Liu, “56-Gb/s/λ C-band DSB IM/DD PAM-4 40-km SSMF transmission employing a multiplier-free MLSE equalizer,” *Optics Express*, 30(7), 11275–11287, 2022.

that “classical” nonlinear equalization can still be adaptive and pilot-efficient (with the usual caution about DD stability)⁷².

7.4.5 Sequence-based detection where relevant (e.g., MLSE/Viterbi as a reference point)

MLSE as a near-optimum ISI benchmark.

For a finite-memory ISI channel, maximum-likelihood sequence estimation (MLSE) selects the symbol sequence $\hat{a}$ maximizing $p(r | a)$. Implemented via the Viterbi algorithm, the trellis complexity grows as M^L for M-ary PAM and memory length L, which is why MLSE is often treated as a performance reference rather than a default solution at very high rates⁷³.

Reduced-state and complexity-curtailed MLSE for IM/DD PAM.

To make MLSE practical in bandwidth-limited IM/DD, reduced-state and multiplier-free variants are actively explored. A reduced-state MLSE for IM/DD PAM is explicitly reported to reduce complexity while retaining much of MLSE’s ISI-mitigation capability⁷⁴. In a 56-Gb/s/λ C-band DSB IM/DD PAM4 demonstration over 40-km SSMF, a multiplier-free MLSE equalizer is reported, showing that careful metric design and implementation choices can bring trellis detection closer to hardware feasibility⁷⁵. A complementary approach proposes a compact look-up-table-based MLSE equalizer for C-band DSB IM/DD

⁷² J. Zhang, H. Tan, X. Hong, J. Liu, C. Guo, C. Fei, X. Wu, A. Lau, S. Yu, and C. Lu, “Comparison of low-complexity sparse and weight-sharing nonlinear equalizers for C-band 100-Gbit/s DSB PAM-4 transmission over 60-km SSMF,” *Optics Express*, 30(20), 36343–36357, 2022.

⁷³ J. G. Proakis and M. Salehi, “*Digital Communications*,” 5th ed., McGraw-Hill, 2007.

⁷⁴ Z. Chen, X. Dai, J. Nie, S. Zhang, J. Zhou, J. Zhang, Y. Qiu, M. Luo, Q. Yang, L. Deng, M. Cheng, K. Qiu, and D. Liu, “Complexity curtailed MLSE equalizer based on a compact look-up-table for C-band DSB IM/DD transmission,” *Optics Letters*, 47(21), 5606–5609, 2022.

⁷⁵ J. Zhou, J. Zhang, X. Zhao, T. Jin, S. Hu, H. Lin, Y. Zhenming, Q. Zhang, B. Xu, and K. Qiu, “Advanced nonlinearity equalizer with TC-NL-MLSE for transmitting beyond 200-Gb/s PAM-8 in IM/DD systems,” *Optics Express*, 30(21), 37416–37425, 2022.

transmission, further emphasizing complexity curtailment through LUT-based metrics⁷⁶.

Nonlinear-aware trellis detection (beyond linear ISI).

When nonlinear effects matter, trellis metrics can be augmented accordingly. An advanced nonlinearity equalizer with TC-NL-MLSE is reported for the transmission of beyond-200-Gb/s PAM8 in IM/DD systems, reflecting the trend toward combining nonlinear modeling with sequence detection under severe ISI⁷⁷. For O-band IM-DD, advanced MLSE methods are also reported in a high-capacity transmission study, underscoring that MLSE variants remain a central benchmark when channel memory is the dominant obstacle⁷⁸.

7.5 ML-Based Equalizers for IM-DD (with emphasis on PAM4)

Before examining the specific motivation, this section differentiates ML-based equalization from the traditional model-driven approach discussed in Sec. 7.4. It presents ML as a complementary method for IM/DD PAM4 links where the effective channel is difficult to model with a single, simple analytical expression. In bandwidth-limited short-reach systems, factors such as memory (inter-symbol interference, ISI), nonlinear distortions (square-law detection and device imperfections), and variations in operating points can make fixed-form equalizers either too complicated (when highly accurate) or too fragile (when simplified). We therefore introduce ML equalizers as adaptable function approximators that learn the necessary inverse mapping directly from data. At a high level, we outline the

⁷⁶ H. Taniguchi, S. Yamamoto, M. Nakamura, A. Masuda, Y. Kisaka, S. Kanazawa, and H. Nakamura, “High-Capacity IM-DD Transmission in O-Band Using Advanced Maximum Likelihood Sequence Estimation Methods,” *Journal of Lightwave Technology*, 41, 3372–3381, 2023.

⁷⁷ T. Xu, Z. Li, J. Peng, A. Tan, Y. Song, Y. Li, J. Chen, and M. Wang, “Decoding of 10-G Optics-Based 50-Gb/s PAM-4 Signal Using Simplified MLSE,” *IEEE Photonics Journal*, 10, 7905508, 2018.

⁷⁸ M. Xiang, S. Fu, O. Xu, J. Li, D. Peng, Z. Gao, Y. Wang, and Y. Qin, “Advanced DSP Enabled C-Band 112 Gbit/s/λ PAM-4 Transmissions With Severe Bandwidth-Constraint,” *Journal of Lightwave Technology*, 40, 987–996, 2022.

main model types, the typical training and deployment process, and what practical “success” means in terms of robustness and implementation constraints.

7.5.1 Why ML helps (nonlinearities + memory + imperfect modeling)

Short-reach IM/DD PAM4 links are often limited by a combination of bandwidth-constrained linear ISI and nonlinear, pattern-dependent distortions. Nonlinearity can arise from square-law direct detection interacting with dispersion, as well as from practical device transfer characteristics (such as modulator/laser chirp behavior, photodiode/TIA nonlinearity, and bandwidth/impedance limits). In these situations, the effective channel exhibits long memory and is hard to model accurately with a simple analytic approach across all operating points. A main reason for using ML equalizers is therefore to reduce the need for explicit models: instead of relying on a single fixed parametric model, the equalizer learns an input–output inverse mapping directly from measured data, thereby implicitly capturing residual nonidealities and effects not accounted for in traditional models⁷⁹.

ML becomes especially appealing when (i) nonlinear ISI is significant, (ii) the channel memory is too extensive for near-optimal sequence methods without aggressive complexity reduction, or (iii) the channel varies enough (temperature, bias drift, component aging, operating point changes) that a “once-calibrated” model is fragile. Recent IM/DD studies demonstrate that NN-based equalization can maintain performance under severe bandwidth constraints by learning nonlinear compensation alongside memory management; for example, neural equalizers that incorporate partial-response structure are specifically designed to use the channel’s limited memory more efficiently than purely memoryless nonlinear compensation⁸⁰. Additionally, comparisons between data-driven neural approaches

⁷⁹ Z. Xu, T. Ji, Q. Wu, W. Lu, H. Ji, Y. Yang, G. Qiao, J. Tang, C. Cheng, L. Liu, S. Wang, J. Liang, Z. Sun, L. Fan, J. Wei, and W. Shieh, “*Advanced Neural Network-Based Equalization in Intensity-Modulated Direct-Detection Optical Systems: Current Status and Future Trends*,” *Photonics*, 11(8), 702, 2024.

⁸⁰ A. Yan, S. Xing, Z. Li, G. Li, A. Sun, P. Luo, J. Shi, H. Zhang, D. Wang, X. Xiao, Z. He, N. Chi, and J. Zhang, “*Adaptive Partial-Response Neural Network Equalization for Bandwidth-Limited*

and structured nonlinear expansions (such as Volterra-type methods used in short-reach optical receivers) demonstrate why ML can be competitive: the network can approximate nonlinear decision boundaries and soft-information mappings without explicitly calculating high-order kernels, whose complexity increases rapidly with memory length⁸¹.

7.5.2 Main families: DNN/MLP, CNN, RNN/LSTM/GRU, reservoir/ESN (high-level comparison)

In IM/DD PAM4 equalization, four architecture families are most common:

- **DNN/MLP (feed-forward networks):** A baseline choice for nonlinear post-equalization using a sliding input window (samples/symbols around the decision point). They are simple to train and map well to parallel hardware, but they represent memory only via the explicit input window length (which increases input dimensionality and computational cost)⁸².
- **1-D CNNs:** Convolutions provide parameter sharing across time and can be efficient for local memory patterns; they often offer a favorable performance/complexity trade-off when the dominant memory is moderate and stationary. CNNs are also highly parallelizable at inference⁸³.
- **RNN / LSTM / GRU:** Designed to capture sequential dependence, they are well-suited when long temporal correlations matter, but their inference is typically less parallel and can be more costly. Complex-valued recurrent equalizers have been demonstrated for PAM4 IM/DD links, showing that

PAM Transmission in Intra-Datacenter Interconnect,” Journal of Lightwave Technology, vol. 42, pp. 2762–2773, 2024.

⁸¹ M. Schädler, G. Böcherer, and S. Pachnicke, “*Soft-Demapping for Short Reach Optical Communication: A Comparison of Deep Neural Networks and Volterra Series*,” Journal of Lightwave Technology, vol. 39, pp. 3095–3105, 2021.

⁸² Z. Xu, T. Ji, Q. Wu, W. Lu, H. Ji, Y. Yang, G. Qiao, J. Tang, C. Cheng, L. Liu, S. Wang, J. Liang, Z. Sun, L. Fan, J. Wei, and W. Shieh, “*Advanced Neural Network-Based Equalization in Intensity-Modulated Direct-Detection Optical Systems: Current Status and Future Trends*,” Photonics, 11(8), 702, 2024.

⁸³ *Ibidem*.

recurrence can be leveraged to track channel memory while reducing complexity⁸⁴.

- **Reservoir computing / ESN:** Reservoir methods provide “cheap” memory by training only a linear readout while keeping the recurrent reservoir fixed. This can drastically simplify training and enable strong performance in short-reach optical scenarios, especially when combined with appropriate pre-processing⁸⁵.

A practical reading is that MLP/CNN architectures often dominate when implementation efficiency and parallel inference are priorities, while RNN-type models are attractive when memory is long or dynamics are important; reservoir **computing** sits between them, offering recurrent memory with a lighter training burden⁸⁶.

7.5.3 Training/inference workflow for equalization (labels, loss, generalization, deployment constraints)

Most ML equalizers for IM/DD PAM4 follow a supervised pipeline:

1. **Data capture and labeling:** Acquire received waveform samples (or DSP features) and align them with known transmitted symbols/bits. Training-data design and validation are nontrivial in optics; a key point is that the dataset must be representative of the operating distribution and should be constructed/validated to avoid misleading generalization (e.g., sequence artifacts, insufficient coverage of distortions)⁸⁷.

⁸⁴ W. Sheng, C. Liu, J. Xiao, L. Sun, Y. Cai, H. Fu, Q. Li, and G. Liu, “Complex-valued recurrent neural network equalizer with low complexity for a 120-Gbps 50-km optical PAM-4 IM/DD system,” *Optics Express*, vol. 32, no. 16, pp. 27624–27634, 2024,

⁸⁵ F. Da Ros, S. Ranzini, H. Bülow, and D. Zibar, “Reservoir-Computing Based Equalization With Optical Pre-Processing for Short-Reach Optical Transmission,” *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 26, pp. 1–12, 2020.

⁸⁶ Z. Xu, T. Ji, Q. Wu, W. Lu, H. Ji, Y. Yang, G. Qiao, J. Tang, C. Cheng, L. Liu, S. Wang, J. Liang, Z. Sun, L. Fan, J. Wei, and W. Shieh, “Advanced Neural Network-Based Equalization in Intensity-Modulated Direct-Detection Optical Systems: Current Status and Future Trends,” *Photonics*, 11(8), 702, 2024.

⁸⁷ T. Liao, L. Xue, L. Huang, W. Hu, and L. Yi, “Training data generation and validation for a neural network-based equalizer,” *Optics Letters*, vol. 45, no. 18, pp. 5113–5116, 2020.

2. **Input representation:** For channels with memory, the standard approach is a windowed mapping (a sliding context of neighboring samples/symbols into the current symbol).
3. **Objective/loss:** Common choices are MSE (regression to symbol levels or soft values) or cross-entropy (classification to PAM levels/bits), depending on whether the equalizer is meant to output hard decisions, soft information, or a denoised continuous estimate.
4. **Complexity control:** Hardware constraints often dominate feasibility. A representative strategy is multi-symbol output (MSO), where a single forward pass predicts multiple consecutive symbols, reducing the number of window evaluations and lowering operations per recovered symbol while maintaining performance in high-rate IM/DD links⁸⁸.
5. **Deployment optimization:** Even when a network achieves a strong BER, its practicality depends on inference cost, latency, and memory footprint. Recent work systematizes complexity accounting and optimization of NN-based equalizers in optical DSP, emphasizing that architectural choices must be co-designed with computational budgets early in the receiver design⁸⁹.

Overall, the “equalizer lifecycle” is best treated as an engineering loop: choose the architecture and context length from channel memory, train with a dataset that reflects expected operating variability, and then compress/optimize the inference path until it meets real-time constraints.

7.5.4 What “success” means (robustness across OSNR, drift, bandwidth, bias, device variations)

For IM/DD PAM4 receivers, *success* cannot be defined by a single BER value at a single OSNR. First, modern optical links are typically evaluated relative to FEC

⁸⁸ B. Sang, W. Zhou, Y. Tan, M. Kong, C. Wang, M. Wang, L. Zhao, J. Zhang, and J. Yu, “Low Complexity Neural Network Equalization Based on Multi-Symbol Output Technique for 200+ Gbps IM/DD Short Reach Optical System,” *Journal of Lightwave Technology*, vol. 40, pp. 2890–2900, 2022.

⁸⁹ P. Freire, S. Srivallapanondh, B. Spinnler, A. Napoli, N. Costa, J. Prilepsky, and S. Turitsyn, “Computational Complexity Optimization of Neural Network-Based Equalizers in Digital Signal Processing: A Comprehensive Approach,” *Journal of Lightwave Technology*, vol. 42, pp. 4177–4201, 2024.

operating points, and pre-FEC BER alone can be a poor predictor of post-FEC behavior, especially when soft-decision FEC and shaping are involved. Information-theoretic metrics such as those used for soft-decision FEC benchmarking provide a more reproducible and system-relevant basis for comparing equalizers across operating points⁹⁰. In particular, post-FEC benchmarking frameworks for bit-interleaved coded modulation (including shaped systems) formalize how to compare designs under equivalent decoding conditions, which is essential when ML equalizers alter the distribution of decision reliabilities⁹¹.

Second, an ML equalizer should be evaluated for robustness: performance stability under varying OSNR, changes in bandwidth limitations (e.g., component replacement), drift in the modulator bias point, and differences across devices. A systematic way to approach this is as a distribution-shift problem: training and deployment environments differ due to drift and hardware variability. While optical IM/DD papers often report OSNR sweeps, the broader ML literature on domain generalization under distribution shifts offers specific concepts and evaluation protocols that can be adapted to establish stronger “robustness” criteria for optical ML equalizers⁹².

7.6 DSP Strategy for BPAM

Section 7.6 introduces bipolar pulse-amplitude modulation (BPAM) with direct detection as a practical way to recover some of the energy efficiency of bipolar signaling while maintaining the hardware simplicity of IM/DD links. Unlike traditional unipolar PAM, BPAM is not directly compatible with square-law photodetection, so the key idea behind modern BPAM/DD is to use oversampling (usually two samples per symbol) and a specially designed transmitter/receiver

⁹⁰ T. Yoshida, M. Karlsson, and E. Agrell, “Performance Metrics for Systems With Soft-Decision FEC and Probabilistic Shaping,” *IEEE Photonics Technology Letters*, vol. 29, pp. 2111–2114, 2017.

⁹¹ T. Yoshida, A. Alvarado, M. Karlsson, and E. Agrell, “Post-FEC BER Benchmarking for Bit-Interleaved Coded Modulation With Probabilistic Shaping,” *Journal of Lightwave Technology*, vol. 38, no. 16, pp. 4292–4306, 2020.

⁹² M. Akrouf, A. Feriani, F. Bellili, A. Mezghani, and E. Hossain, “Domain Generalization in Machine Learning Models for Wireless Communications: Concepts, State-of-the-Art, and Open Issues,” *IEEE Communications Surveys & Tutorials*, vol. 25, pp. 3014–3037, 2023.

processing chain that enables the receiver to recover the information otherwise “lost” in intensity detection. The goal of this section is twofold: first, to give a current overview of oversampling-enabled BPAM/DD; and second, to set the stage for the following subsections, where BPAM-specific DSP and equalization (including our experimental CNN-based equalizer) are discussed in detail⁹³.

7.6.1 BPAM: State of the art of oversampling-enabled BPAM/DD

BPAM is a well-known concept in digital communications, where amplitude levels can be both positive and negative. However, in optical IM/DD systems, direct detection is inherently square-law, meaning that the photodiode current is proportional to the optical intensity. This characteristic removes the sign or phase information that bipolar signaling usually depends on. As a result, short-range optical systems have traditionally used unipolar formats, even though bipolar formats are, in theory, more energy-efficient⁹⁴.

A major breakthrough was the introduction of a BPAM/DD receiver strategy that uses oversampling at two samples per symbol to gather a sufficient set of observables from the detected waveform. This allows bipolar detection with a single photodetector and requires no additional bandwidth beyond what a typical system already needs. In this method, the receiver leverages the extra time resolution instead of additional coherent hardware and applies a simple, structured post-processing step to reconstruct the BPAM symbol sequence.⁹⁵

Beyond the core principle, early experimental validation showed that BPAM/DD can provide significant system-level improvements in realistic short-reach scenarios. Specifically, the BPAM/DD technique reported by Secondini et al. demonstrated an OSNR gain of about 4.5 dB compared to conventional IM/DD

⁹³ M. Secondini and E. Forestieri, “Direct Detection of Bipolar Pulse Amplitude Modulation”, *Journal of Lightwave Technology*, vol. 38, no. 21, pp. 5981–5990, 2020.

⁹⁴ *Ibidem*.

⁹⁵ *Ibidem*.

baselines and exhibited good tolerance to group-velocity dispersion, while only requiring a small increase in receiver-side processing complexity⁹⁶.

Following these enabling demonstrations, the BPAM/DD literature has increasingly emphasized information-theoretic and system-level comparisons between direct detection and coherent detection, as well as the role of oversampling in approaching the achievable-rate performance of more complex receivers. In this line of work, achievable-rate studies for short-reach fiber-optic channels with direct detection provide a formal framework for understanding when (and how far) oversampled DD architectures can close the gap to coherent reception under realistic noise and bandwidth constraints⁹⁷.

A parallel, complementary direction is the broader post-200G IM/DD ecosystem, where modulation format options and DSP complexity are discussed in relation to bandwidth constraints, implementation practicality, and standardization pressures. In this environment, BPAM/DD is naturally positioned as an “advanced IM/DD” option: it seeks to utilize additional degrees of freedom through sampling and DSP rather than a full coherent front-end, but its adoption relies on whether its DSP load and robustness can compete with the rapidly progressing unipolar PAM toolbox⁹⁸.

Finally, recent IM/DD research trends increasingly focus on data-driven receiver optimization, including learned transceiver mappings and ML-assisted compensation for channel and hardware impairments. Although much of this work is not BPAM-specific, it is directly relevant because BPAM/DD receivers deliberately depend on DSP to recover information obscured by square-law detection, making them natural candidates for ML-based equalization and detection

⁹⁶ M. Secondini, E. Forestieri, F. Fresi, L. Potì, T. Catuogno, and F. Cavaliere, “*Bipolar Pulse Amplitude Modulation with Direct Detection*”, in 45th European Conference on Optical Communication (ECOC 2019), 2019.

⁹⁷ D. Plabst, T. Prinz, T. Wiegart, S. Calabrò, T. Rahman, N. Stojanović, N. Hanik, and G. Kramer, “*Achievable Rates for Short-Reach Fiber-Optic Channels With Direct Detection*”, *Journal of Lightwave Technology*, vol. 40, pp. 3602–3613, 2021.

⁹⁸ D. Che and X. Chen, “*Modulation Format and Digital Signal Processing for IM-DD Optics at Post-200G Era*”, *Journal of Lightwave Technology*, vol. 42, pp. 588–605, 2024.

when deterministic models are inadequate, or hardware distortions dominate⁹⁹. In this thesis, we build upon these developments by experimentally evaluating BPAM/DD with a CNN equalizer, comparing it to an adaptive equalizer benchmark from the BPAM/DD experimental literature, and then using the results to motivate the BPAM-specific DSP strategy described in the following subsections¹⁰⁰.

7.6.2 Analytical basis of oversampling-enabled BPAM/DD

A conventional IM-DD receiver measures optical intensity (square-law), which, if interpreted locally, seems to discard the sign of a bipolar symbol. The important insight behind BPAM with direct detection is that square-law detection not only creates “self-terms” (which erase the sign) but also produces cross-terms between overlapping symbols. These cross-terms preserve the relative sign between adjacent bipolar symbols. By intentionally sampling the detected waveform at two points per symbol ($T/2$ -spaced) and selecting a pulse shape or receive filtering that makes nearest-neighbor overlap dominant at half-symbol instants, you get a pair of discrete-time observations per symbol: one that is primarily “amplitude-like” (mainly representing a single symbol) and another that is mainly “product-like” (representing the product of two adjacent symbols). This mechanism enables bipolar recovery with a single square-law detector photodiode¹⁰¹.

Discrete-time model (2 samples/symbol, square-law DD)

Consider bipolar 2-PAM (BPAM) symbols $a_k \in \{+1, -1\}$, symbol period T , and a real Nyquist pulse $p(t)$. The transmitted (real) baseband optical field is modeled as

⁹⁹ A. Omid, M. Banawan, E. Weckenmann, B. Paquin, A. Geravand, Z. Zheng, W. Shi, M. Zeng, and L. Rusch, “Auto-Encoder Optimized PAM IM/DD Transceivers for Amplified Fiber Links”, *Journal of Lightwave Technology*, vol. 43, pp. 547–556, 2024.

¹⁰⁰ R. Solaimani, A. Nizamani, S. Civelli, P. Nadimi Goki, F. Cavaliere, and L. Poti, “Experimental Demonstration of Convolutional Neural Network Equalization for BPAM in IM-DD Systems”, *Advanced Photonics Congress (IPR, Networks, NOMA, SOLITH, SPPCom)* 2025.

¹⁰¹ M. Secondini and E. Forestieri, “Direct Detection of Bipolar Pulse Amplitude Modulation”, *Journal of Lightwave Technology*, vol. 38, no. 21, pp. 5981–5990, 2020.

$$s(t) = \sum_k a_k p(t - kT).$$

Ignoring noise for the moment, the direct-detection output is proportional to the squared magnitude of the field. With a real field model, this reduces to

$$y(t) = |s(t)|^2 = s^2(t).$$

Expanding the square yields a sum of self-terms and cross-terms:

$$\begin{aligned} y(t) &= \left(\sum_k a_k p(t - kT) \right)^2 \\ &= \sum_k a_k^2 p^2(t - kT) + 2 \sum_{k < \ell} a_k a_\ell p(t - kT) p(t - \ell T). \end{aligned}$$

For bipolar symbols, $a_k^2 = 1$, so the self-term $\sum_k a_k^2 p^2(\cdot)$ does *not* reveal the sign of a_k . The sign information survives only through the cross-terms ($a_k a_\ell$), which encode *relative sign* between symbols that overlap in time.

Now sample at twice the symbol rate: $t = nT/2$, and define $y_n = y(nT/2)$. For a properly chosen pulse shape and/or a receive filter bank aligned with the two sampling phases, the two phases can be arranged so that (i) the symbol-center samples are dominated by a single-symbol contribution, whereas (ii) the half-symbol-offset samples are dominated by the nearest-neighbor overlap term. This yields an approximate two-branch discrete model of the form

$$r_k = \begin{bmatrix} r_{k,0} \\ r_{k,1} \end{bmatrix} \approx \begin{bmatrix} \alpha_0 + \beta_0 a_k \\ \alpha_1 + \beta_1 a_k a_{k+1} \end{bmatrix} + \begin{bmatrix} w_{k,0} \\ w_{r_{k,1}} \end{bmatrix},$$

where α_i, β_i are constants determined by the pulse shape, scaling, and filtering, and $w_{k,i}$ represent the (aggregate) noise contributions after sampling. The first component $r_{k,0}$ is “amplitude-like” (single-symbol dominated), while the second

component $r_{k,1}$ is “product-like,” encoding the adjacent-symbol product $a_k a_{k+1}$.

102

Key cross-term intuition (why bipolar recovery becomes possible)

- **Local sign loss:** because $a_k^2 = 1$, any self-term depending only on a_k^2 cannot distinguish +1 from -1.
- **Relative sign retention:** the cross-term between adjacent symbols contains $a_k a_{k+1}$, which equals +1 if the two symbols have the same sign and -1 if they have opposite signs. Thus, while the detector cannot “see” the sign of a single isolated symbol, it can “see” whether two neighbors have equal or opposite signs through the overlap contribution.
- **Why 2× oversampling matters:** with two samples per symbol (or an equivalent two-phase filter bank), each symbol contributes to (i) a sample that is primarily tied to that symbol and (ii) samples that are primarily tied to *products* with its neighbors. Jointly, the sequence of “self-like” and “product-like” observations provides enough structure to reconstruct the bipolar sign pattern¹⁰³.

A clean conceptual detection rule

Assume symbol timing is known and the receiver provides the two statistics $\{r_{k,0}, r_{k,1}\}$. A simple conceptual recovery strategy is:

1. Center (self-like) decision:

$$\hat{a}_k^0 = \text{sign}(r_{k,0} - \theta_0),$$

where θ_0 is a threshold centered around α_0 .

2. Midpoint (product-like) decision:

$$(\widehat{a_k a_{k+1}}) = \text{sign}(r_{k,1} - \theta_1),$$

where θ_1 is a threshold centered around α_1 .

3. Consistency propagation (sequence reconstruction):

starting from an initial decision $\hat{a}_0$, propagate

¹⁰² *Ibidem*.

¹⁰³ *Ibidem*.

$$\hat{a}_{k+1} \leftarrow \text{sign}((\widehat{a_k a_{k+1}}) \cdot \widehat{a_k}),$$

optionally fusing this propagated estimate with the direct center estimate $\hat{a}_{k+1}^0$.

In practice, because noise and residual ISI can make hard propagation fragile, the same underlying model naturally motivates *maximum-likelihood (ML) sequence detection* (or equivalent trellis-based inference) that jointly exploits the whole set of observations $\{r_k\}$ across time rather than committing to early hard decisions. The key point is that the detection metric must leverage both the “self-like” and “product-like” branches, as the latter encodes bipolar sign relations¹⁰⁴.

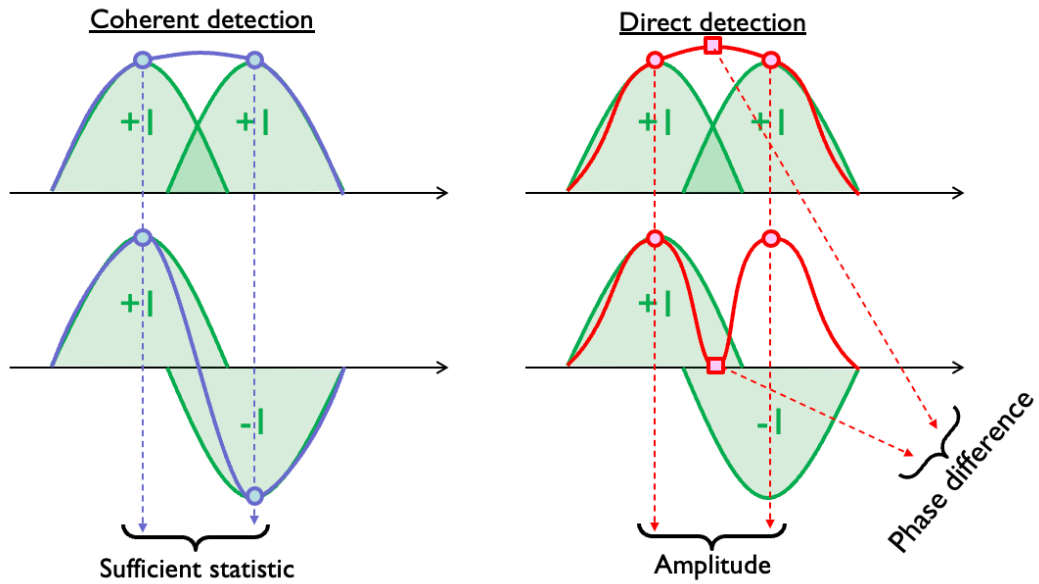


Figure 7.1 Basic principle for detection of amplitude and phase by oversampling¹⁰⁵

7.6.3 BPAM-specific processing chain and experimental requirements

From a DSP perspective, a BPAM direct-detection receiver shares many “plumbing” steps with a conventional PAM4 IM-DD receiver (data capture, front-end filtering, timing alignment, equalization, slicing, and BER evaluation). However, BPAM requires a key architectural change: the receiver must preserve and utilize fractionally spaced information (most commonly two samples per

¹⁰⁴ *Ibidem.*

¹⁰⁵ *Ibidem.*

symbol, $T/2$) because the bipolar information cannot be recovered from symbol-rate intensity samples alone. In oversampling-enabled BPAM/DD, the $T/2$ -spaced samples provide a structured set of observables where (i) one sampling phase is mainly “self-like” and (ii) the other sampling phase is mainly “cross-term-like,” allowing for the reconstruction of the bipolar sign pattern after square-law detection¹⁰⁶.

What must change vs. standard PAM4 DSP (BPAM-specific deltas)

- **Mandatory fractionally spaced processing:** while PAM4 systems may downsample early to symbol rate once timing is established, BPAM/DD must preserve a fractionally spaced representation (typically 2 samples/symbol) through the equalization/detection stages to retain the cross-term information that encodes sign relations¹⁰⁷.
- **Two-part decision logic (sign + magnitude):** for bipolar alphabets, the decision stage naturally separates into (i) sign recovery (using cross-term-bearing observables across adjacent symbols) and (ii) magnitude decision/slicing (when multi-level BPAM is used). The sign-recovery component is absent in standard unipolar PAM4 DD slicing¹⁰⁸.
- **Equalizer operating point:** equalization is most naturally implemented as a *fractionally spaced FFE* (and, if needed, a nonlinear extension) so that timing-phase sensitivity and bandwidth/dispersion-induced memory are mitigated without prematurely collapsing the two-phase structure¹⁰⁹.

Practical BPAM/DD receiver DSP chain (offline processing)

(1) Data acquisition and initial resampling.

¹⁰⁶ *Ibidem*.

¹⁰⁷ *Ibidem*.

¹⁰⁸ *Ibidem*.

¹⁰⁹ X. Wu, A. Karar, K. Zhong, A. Lau, and C. Lu, “Experimental demonstration of pre-electronic dispersion compensation in IM/DD systems using an iterative algorithm”, *Optics Express*, vol. 29, no. 16, pp. 24735–24749, 2021.

In experimental IM-DD links, the photodiode/TIA output is typically digitized at a high sampling rate (multiple samples per symbol) to support robust offline synchronization and DSP, after which the waveform is numerically resampled to a convenient fractionally spaced rate (commonly two samples/symbol) used by timing/equalization blocks¹¹⁰.

(2) Front-end digital filtering (noise/bandwidth conditioning).

A receiver-side digital low-pass filter is commonly applied to suppress out-of-band noise, emulate/compensate the effective front-end electrical bandwidth, and avoid aliasing at the chosen DSP working rate. The importance of bandwidth-aware sampling/processing is especially pronounced for Nyquist-shaped signaling under limited receiver bandwidth, where sampling-assisted strategies can improve detectability¹¹¹.

(3) Timing alignment/synchronization.

Before detection, timing recovery (or equivalent synchronization) is needed to correct sampling-frequency and sampling-phase offsets. In practice, working at a *fractionally spaced rate* and employing a fractionally spaced equalizer provides additional robustness to residual timing-phase errors while simultaneously addressing channel memory¹¹².

(4) Fractionally spaced equalization and BPAM feature construction.

At two samples/symbol, the input to the equalizer is naturally represented as a tapped-delay line of $T/2$ -spaced samples. For BPAM/DD, the equalizer output (or intermediate statistics derived from it) must preserve the two-phase structure that

¹¹⁰ M. Hossain, G. Böcherer, T. Rahman, T. Wettlin, N. Stojanović, S. Calabrò, and S. Pachnicke, “Probabilistic Shaping for High-Speed Unamplified IM/DD Systems With an O-Band EML”, *Journal of Lightwave Technology*, vol. 41, pp. 5373–5382, 2023.

¹¹¹ Z. Geng, D. Kong, V. Rozental, A. Lowery, and B. Corcoran, “Optical sampling to enhance Nyquist-shaped signal detection under limited receiver bandwidth”, *Optics Express*, vol. 27, no. 17, pp. 24007–24017, 2019.

¹¹² *Ibidem*. X. Wu *et al.*

corresponds to “self-dominated” and “cross-term-dominated” observations; these paired observables are then used as the features for bipolar recovery¹¹³.

(5) Optional nonlinear compensation (when distortions dominate).

When the link is strongly distortion-limited (e.g., bandwidth-limited plus nonlinearity), nonlinear compensation may be required in addition to linear FFE. In IM-DD data-center link analysis, distortion-limited operation is a primary driver for moving beyond purely linear models and motivates strategies that explicitly address nonlinear impairments¹¹⁴.

(6) Downsampling and decision stage (sign-bit + magnitude).

Once timing is locked and equalization converges, the receiver makes symbol decisions. For BPAM, the sign cannot be reliably determined from intensity samples alone; instead, sign reconstruction uses cross-term-bearing statistics across adjacent symbols (enabled by the $T/2$ representation). For multi-level bipolar formats, after recovering the sign, the magnitude can be determined by amplitude slicing (or soft-decision metrics) on an equalized symbol variable, similar to magnitude decisions in multilevel IM-DD receivers¹¹⁵.

(7) Performance evaluation.

As in typical IM-DD experiments, BER is calculated by comparing the decided bitstream to a known transmitted sequence (often PRBS-based) after aligning the recovered data and the reference¹¹⁶.

Experimental requirement checklist

In oversampling-enabled BPAM/DD, the most critical receiver-side requirement is to retain a fractionally spaced representation (typically two samples/symbol, i.e., $T/2$ -spaced) through the equalization and detection stages,

¹¹³ M. Secondini and E. Forestieri, “Direct Detection of Bipolar Pulse Amplitude Modulation”, *Journal of Lightwave Technology*, vol. 38, no. 21, pp. 5981–5990, 2020.

¹¹⁴ E. Liang and J. Kahn, “Geometric Shaping for Distortion-Limited Intensity Modulation/Direct Detection Data Center Links”, *IEEE Photonics Journal*, vol. 15, pp. 1–17, 2023.

¹¹⁵ *Ibidem*. M. Secondini and E. Forestieri.

¹¹⁶ M. Hossain, G. Böcherer, T. Rahman, T. Wettlin, N. Stojanović, S. Calabrò, and S. Pachnicke, “Probabilistic Shaping for High-Speed Unamplified IM/DD Systems With an O-Band EML”, *Journal of Lightwave Technology*, vol. 41, pp. 5373–5382, 2023.

because bipolar recovery relies on observables created by the cross-terms of square-law detection that are not accessible from symbol-rate intensity samples alone. Consequently, the working DSP rate and the internal signal representation must preserve the “two-phase” structure that jointly carries (i) a predominantly self-dominated observable and (ii) a predominantly neighbor-product-dominated observable used for sign reconstruction¹¹⁷.

From an experimental standpoint, BPAM/DD systems are typically digitized with a high acquisition sampling rate (multiple samples per symbol) using a *real-time oscilloscope* or equivalent ADC to support robust offline synchronization and processing. The acquired waveform is then numerically resampled to a convenient fractionally spaced working rate (commonly two samples/symbol) for timing recovery and equalization, and only after convergence is a final symbol-rate stream derived for decisions and BER evaluation. This high-rate capture leads to a controlled resampling approach that is consistent with modern IM/DD experimental DSP practice, especially when the receiver must remain robust to non-ideal front-end responses and clocking offsets¹¹⁸.

A second practical requirement is that the receiver DSP must be *bandwidth-aware*. In Nyquist-like signaling under limited receiver bandwidth, the effective analog front-end response reshapes the waveform. It increases memory/ISI, so digital processing typically includes front-end low-pass filtering (and, where relevant, sampling-assisted strategies) to control out-of-band noise, avoid aliasing at the chosen working rate, and maintain detectability when the receiver bandwidth is a limiting factor. This becomes particularly important when BPAM/DD relies on consistent fractionally spaced statistics, which can be degraded if bandwidth-

¹¹⁷ M. Secondini and E. Forestieri, “Direct Detection of Bipolar Pulse Amplitude Modulation”, *Journal of Lightwave Technology*, vol. 38, no. 21, pp. 5981–5990, 2020.

¹¹⁸ M. Hossain, G. Böcherer, T. Rahman, T. Wettlin, N. Stojanović, S. Calabrò, and S. Pachnicke, “Probabilistic Shaping for High-Speed Unamplified IM/DD Systems With an O-Band EML”, *Journal of Lightwave Technology*, vol. 41, pp. 5373–5382, 2023.

induced distortions blur the mid-symbol overlap structure exploited for bipolar recovery¹¹⁹.

Timing sensitivity is also a central experimental constraint: symbol-rate sampling is highly sensitive to sampling-phase offsets, whereas fractionally spaced equalization at $T/2$ relaxes this sensitivity by providing additional degrees of freedom to compensate for residual timing-phase error jointly with ISI. Moreover, for dispersive links, the BPAM/DD receiver chain must include dispersion-aware processing, such as pre- and/or post-electronic dispersion compensation coupled with fractionally spaced equalization to restore the temporal structure required for stable sign recovery and reliable slicing¹²⁰.

When operating conditions are *distortion-limited* (e.g., strong bandwidth limitation combined with nonlinear transmitter/receiver impairments), linear equalization alone may be insufficient, because nonlinear distortions can corrupt both the amplitude-like and the cross-term-like observables that BPAM/DD uses for reconstruction. In such regimes, the receiver must either employ nonlinear compensation/equalization strategies or otherwise shape the signaling to remain robust under nonlinear memory effects; this requirement is not unique to BPAM, but it becomes more consequential because BPAM/DD detection depends on preserving subtle structure in the fractionally spaced waveform¹²¹.

Finally, if the transceiver applies algorithmic compensation stages that combine *lookup-table (LUT)-based correction* with iterative reconstruction methods (e.g., Gerchberg–Saxton-type processing) for mitigating nonlinearities and chromatic dispersion, the BPAM/DD receiver must verify that these stages do not inadvertently destroy the statistics used for sign reconstruction. In practice,

¹¹⁹ Z. Geng, D. Kong, V. Rozental, A. Lowery, and B. Corcoran, “Optical sampling to enhance Nyquist-shaped signal detection under limited receiver bandwidth”, *Optics Express*, vol. 27, no. 17, pp. 24007–24017, 2019.

¹²⁰ X. Wu, A. Karar, K. Zhong, A. Lau, and C. Lu, “Experimental demonstration of pre-electronic dispersion compensation in IM/DD systems using an iterative algorithm”, *Optics Express*, vol. 29, no. 16, pp. 24735–24749, 2021.

¹²¹ E. Liang and J. Kahn, “Geometric Shaping for Distortion-Limited Intensity Modulation/Direct Detection Data Center Links”, *IEEE Photonics Journal*, vol. 15, pp. 1–17, 2023.

compensation blocks should be assessed not only by BER improvement but also by their ability to preserve (or enhance) the stability of the fractionally spaced BPAM features that feed the equalizer and the final sign/magnitude decision logic¹²².

7.6.4 Equalization implications for BPAM

In oversampling-enabled BPAM with direct detection, equalization is not just a traditional “PAM clean-up” stage placed before a memoryless slicer; it is intrinsically linked to the bipolar recovery mechanism. This is because BPAM/DD depends on a fractionally spaced representation (usually $T/2$) to maintain the two-phase observables produced by square-law detection, where sign-related information is carried by cross-terms instead of symbol-rate intensity samples. Therefore, equalization must be designed to reduce channel memory (bandwidth-limited ISI and dispersion-induced distortion) while preserving the waveform structure necessary for sign reconstruction¹²³. More broadly, in IM/DD links operated at very high symbol rates, the effective channel can exhibit not only linear ISI but also nontrivial distortion mechanisms (e.g., chirp-dispersion interactions and square-law-related nonlinearities), which motivates a hierarchy of equalizer families with increasing modeling power and implementation cost¹²⁴.

The baseline equalizer family most directly aligned with BPAM/DD is the *fractionally spaced linear feed-forward equalizer (FFE)* operating at (2) samples per symbol. In an oversampled BPAM chain, this equalizer is placed immediately after the ADC (or after numerical resampling to the target $T/2$ grid) and before the final bipolar decision logic, because its primary role is to reduce bandwidth-limited ISI and dispersion-induced memory while retaining the $T/2$ -spaced statistics needed by BPAM recovery. A key practical advantage of this choice is that

¹²² X. Zhang, J. Huo, H. Bai, P. Qin, H. Wei, and K. Long, “Performance Investigation of Joint LUT and GS Algorithm at the Transceiver for Nonlinear and CD Compensation”, *Photonics*, 11(7), 665, 2024.

¹²³ M. Secondini and E. Forestieri, “Direct Detection of Bipolar Pulse Amplitude Modulation”, *Journal of Lightwave Technology*, vol. 38, no. 21, pp. 5981–5990, 2020.

¹²⁴ D. Che and X. Chen, “Modulation Format and Digital Signal Processing for IM-DD Optics at Post-200G Era”, *Journal of Lightwave Technology*, vol. 42, pp. 588–605, 2024.

fractionally spaced equalizers also relax sampling-phase sensitivity compared to symbol-spaced processing, which is beneficial in experiments where residual timing offsets and imperfect synchronization are unavoidable¹²⁵. When the receiver front-end exhibits joint optical-electrical behavior that limits dispersion-limited performance, feed-forward equalization can also be formulated to jointly account for optical and electrical effects while remaining within a low-complexity linear framework, making it suitable as an early DSP block before any nonlinear compensation is attempted¹²⁶. Although BPAM demonstrations are often single-channel, it is worth noting that the same fractionally spaced “linear-first” philosophy extends naturally to more complex IM/DD scenarios such as strongly coupled mode-division multiplexing, where a frequency-domain MIMO equalizer can be used as the principal linear equalization engine at the receiver¹²⁷.

Independent of whether the equalizer is realized in the time domain (FFE/DFE) or the frequency domain (FDE/MIMO-FDE), practical implementations generally rely on adaptive coefficient updates. In the IM/DD literature, *LMS-type adaptation* is frequently used because it offers very low complexity and can support long filters and high-dimensional equalizers in offline processing; for example, large-tap equalizers can remain computationally feasible when trained and tracked with LMS-style updates¹²⁸. In principle, faster-converging updates such as RLS can be employed when rapid tracking is required, but this comes at substantially higher computational cost; in a BPAM context, the key requirement is that the adaptation

¹²⁵ X. Wu, A. Karar, K. Zhong, A. Lau, and C. Lu, “Experimental demonstration of pre-electronic dispersion compensation in IM/DD systems using an iterative algorithm”, *Optics Express*, vol. 29, no. 16, pp. 24735–24749, 2021.

¹²⁶ P. Zhu, Y. Yoshida, A. Kanno, and K. Kitayama, “Analysis and Demonstration of Low-Complexity Joint Optical-Electrical Feedforward Equalization (OE-FFE) For Dispersion-Limited High-Speed IM/DD Transmission”, *Journal of Lightwave Technology*, vol. 41, pp. 477–488, 2023.

¹²⁷ J. Long, C. Wang, Y. Wu, B. Sang, C. Bian, X. Yang, L. Zhang, Y. Chen, Q. Zhang, Y. Wang, Y. Li, W. Zhou, K. Wang, B. Liu, L. Shen, and J. Yu, “Digital Frequency-Domain MIMO Equalizer Enabling Six-LP-Mode Strong-Coupling IM/DD MDM Optical Transmission System”, *Sensors*, vol. 25, 2025.

¹²⁸ *Ibidem*.

mechanism must operate on the fractionally spaced stream so that the equalizer converges in the same sample space where bipolar-relevant structure is present.

A second major equalizer family relevant to BPAM benchmarking is the *decision-feedback equalizer (DFE)*, typically applied after an initial linear front-end has reduced precursor ISI and stabilized timing. In oversampled chains, DFEs are commonly run at *symbol rate* on decided symbols and are especially attractive in strongly bandwidth-limited links because they can mitigate post-cursor ISI while avoiding the same degree of noise enhancement that can occur with purely feed-forward structures near spectral notches. However, DFEs introduce the fundamental risk of *error propagation*, and in IM/DD links, this can manifest as bursty error events unless the detection architecture includes additional mechanisms such as soft-decision processing to control feedback-induced error amplification¹²⁹.

When ISI becomes sufficiently severe that symbol-by-symbol slicing (even with DFE) is no longer effective, sequence-based equalization and detection become relevant. In this class, *reduced-state maximum-likelihood sequence estimation (MLSE)* is a common compromise: it seeks to approach the performance of ML detection while limiting complexity by shortening channel memory with a linear front end and restricting the trellis state space. Such approaches are directly applicable to bandwidth-limited IM/DD channels with PAM signaling and are conceptually relevant to BPAM whenever the receiver's effective memory spans multiple symbols and cannot be reliably collapsed into a short feedforward structure without performance loss¹³⁰. A closely related direction combines a front-end equalizer with simplified soft-output MLSE and coding, which is particularly relevant when the receiver must provide soft information to an FEC decoder; in

¹²⁹ K. Wu, G. Liga, J. Riani, and A. Alvarado, "Low-Complexity Soft-Decision Detection for Combating DFE Burst Errors in IM/DD Links", *Journal of Lightwave Technology*, vol. 42, pp. 1395–1408, 2023.

¹³⁰ Y. Yu, Y. Che, T. Bo, D. Kim, and H. Kim, "Reduced-state MLSE for an IM/DD system using PAM modulation", *Optics Express*, vol. 28, no. 26, pp. 38505–38515, 2020.

these architectures, the FFE is explicitly used to make trellis-based detection tractable¹³¹.

As symbol rate and distortion levels increase, the effective IM/DD channel can become strongly nonlinear with memory, motivating *Volterra and memory-polynomial equalizers*. In this family, a nonlinear feed-forward equalizer (often a second-order Volterra structure in its simplest practical form) is used to capture joint linear and nonlinear distortions such as chirp-dispersion beating and square-law-induced nonlinear memory. A central advantage is modeling power: even relatively low-order Volterra equalizers can substantially improve performance in directly modulated laser (DML) based IM/DD systems, where purely linear equalization may saturate¹³². The main limitation is complexity: the number of coefficients grows rapidly with memory length and polynomial order, so practical designs frequently adopt sparsity, pruning, or structured coefficient sharing to reduce multiplications while retaining most of the performance gain. In this context, sparse and weight-sharing nonlinear equalizer families have been compared specifically to quantify the performance-complexity trade-off in high-speed IM/DD PAM links¹³³. A particularly effective simplification strategy is to retain only selected nonlinear terms that dominate the impairment physics; for example, absolute-term-based nonlinear equalizers represent an explicit attempt to capture dominant nonlinear contributions with substantially reduced kernel sets compared to full Volterra expansions¹³⁴.

¹³¹ X. Zhao, J. Zhang, J. Zhou, S. Hu, B. Xu, Q. Zhang, and K. Qiu, “Collaborative FFE-assisted simplified soft-output MLSE with LDPC for high-speed IM-DD transmissions”, *Optics Express*, vol. 32, no. 3, pp. 3866–3873, 2024.

¹³² Y. Yu, M. Choi, T. Bo, Z. He, Y. Che, and H. Kim, “Low-Complexity Second-Order Volterra Equalizer for DML-Based IM/DD Transmission System”, *Journal of Lightwave Technology*, vol. 38, pp. 1735–1746, 2020.

¹³³ J. Zhang, H. Tan, X. Hong, J. Liu, C. Guo, C. Fei, X. Wu, A. Lau, S. Yu, and C. Lu, “Comparison of low-complexity sparse and weight-sharing nonlinear equalizers for C-band 100-Gbit/s DSB PAM-4 transmission over 60-km SSMF”, *Optics Express*, vol. 30, no. 20, pp. 36343–36357, 2022.

¹³⁴ J. Zhang, Z. Lin, X. Wu, J. Liu, A. Lau, C. Guo, C. Lu, and S. Yu, “Low-complexity sparse absolute-term based nonlinear equalizer for C-band IM/DD systems”, *Optics Express*, vol. 29, no. 14, pp. 21891–21901, 2021.

Nonlinear compensation can also be placed in *decision-feedback* form. Volterra-DFE and memory-polynomial DFE variants are motivated by the observation that feedback structures can emulate IIR-like behavior and can therefore equalize channels with deep fading notches and strong memory more efficiently than a purely feed-forward nonlinear structure of comparable complexity. However, the error-propagation vulnerability of DFEs becomes more severe in nonlinear feedback architectures, and practical implementations must control complexity through kernel simplification, structured bases, or LUT-assisted methods. One example is the trigonometric-memory-polynomial DFE, which is explicitly introduced as a low-complexity nonlinear compensation approach for IM/DD optical systems¹³⁵. Another example is simplified Volterra-structure DFE formulations that target nonlinear compensation in PAM4 IM-DD systems while keeping computational cost manageable¹³⁶. In addition, LUT-assisted nonlinear DFE variants have been proposed to further reduce complexity while preserving the core benefit of nonlinear decision feedback in IM/DD links¹³⁷. For BPAM, the key implication is architectural: nonlinear equalization must still be integrated so that it does not destroy the fractionally spaced structure that the BPAM bipolar decision relies on; hence, nonlinear DFEs are typically preceded by a stabilizing linear front-end at the oversampled rate.

Finally, ML-based equalizers provide an alternative nonlinear-compensation route when the impairment landscape is strongly nonlinear, partially stochastic, or poorly captured by tractable analytical models. In IM/DD research, lightweight deep neural equalizers have been demonstrated in challenging links (including high-frequency photonic wireless IM/DD transmission) and are often deployed as

¹³⁵ X. Liu, Y. Liu, C. Hu, and J. Zhao, “*Experimental demonstration of low-complexity trigonometric-memory-polynomial decision-feedback equalizer for nonlinear compensation in IM/DD optical systems*”, *Optics Express*, vol. 32, no. 3, pp. 3329–3341, 2024.

¹³⁶ Y. Xie, P. He, W. Li, and N. Li, “*A Decision Feedback Equalization Algorithm Based on Simplified Volterra Structure for PAM4 IM-DD Optical Communication Systems*”, *Applied Sciences*, 2023.

¹³⁷ J. Zhang, H. Tan, A. Lau, Z. Li, and C. Lu, “*Low-complexity cluster-assisting look-up-table-based Volterra decision-feedback equalizer for IM/DD systems*”, *Optics Letters*, vol. 50, no. 3, pp. 992–995, 2025.

a nonlinear stage after basic synchronization, frequently in combination with auxiliary clustering or decision-aiding mechanisms to improve robustness¹³⁸. Two-stage deep-learning equalization has also been proposed to explicitly improve nonlinearity robustness in IM/DD THz systems, reflecting the broader trend of using staged ML pipelines (front-end conditioning followed by nonlinear inference) when single-stage equalizers are insufficient or inefficient¹³⁹. Beyond standard DNNs, alternative ML structures such as adaptive-network-based fuzzy inference system equalizers and Kolmogorov-Arnold network equalizers have been proposed to achieve favorable performance-complexity trade-offs in IM/DD settings¹⁴⁰. The general implication for BPAM is that ML equalizers can be viewed as a flexible, data-driven mechanism to learn the nonlinear mapping from oversampled intensity waveforms (or equalizer-conditioned features) to bipolar decisions, provided that training and generalization are carefully controlled.

From the standpoint of benchmark framing in this thesis, the most defensible “classical” reference receiver for oversampled BPAM/DD is therefore a *fractionally spaced linear FFE operating at $T/2$* , trained with a standard low-complexity adaptive procedure, and used as the front-end that preserves the BPAM-relevant two-phase structure before final decisions. This benchmark is technically justified because it is the simplest equalization architecture that remains consistent with the BPAM requirement of exploiting oversampled observables, while also

¹³⁸ J. Ge, J. Zhang, S. Xu, Q. Wang, J. Lin, S. Hu, X. Lu, Z. Ou, S. Wang, T. Wang, Y. Li, J. Chen, T. Zhang, and W. Zhou, “82.5 GHz Photonic W-Band IM/DD PS-PAM4 Wireless Transmission over 300 m Based on Balanced and Lightweight DNN Equalizer Cascaded with Clustering Algorithm”, *Sensors*, vol. 25, 2025.

¹³⁹ P. Torkaman, S. Latifi, K. Feng, and S. Yang, “Nonlinearity-Robust IM/DD THz Communication System via Two-Stage Deep Learning Equalizer”, *IEEE Communications Letters*, vol. 28, pp. 1805–1809, 2024.

¹⁴⁰ F. Wang, R. Gao, Z. Li, J. Liu, Y. Cui, Q. Xu, X. Pan, L. Zhu, D. Guo, H. Chang, S. Zhou, Z. Dong, Q. Zhang, Q. Tian, F. Tian, X. Huang, J. Yan, L. Jiang, and X. Xin, “OAM mode-division multiplexing IM/DD transmission at 4.32 Tbit/s with a low-complexity adaptive-network-based fuzzy inference system nonlinear equalizer”, *Optics Letters*, vol. 49, no. 3, pp. 430–433, 2023.

reflecting the dominant equalization practice in experimental IM/DD links¹⁴¹. More advanced equalizers, such as nonlinear Volterra/memory-polynomial families, reduced-state sequence detection, or ML-based models, are then naturally motivated as complexity-increasing options when the system is limited by nonlinear memory or by severe residual ISI that cannot be sufficiently mitigated by a linear fractionally spaced front-end alone¹⁴².

7.7 Experimental Demonstration of Convolutional Neural Network Equalization for BPAM in IM-DD Systems

The previous section established why BPAM can be recovered under square-law direct detection by deliberately exploiting oversampling, and it clarified the DSP implications that distinguish BPAM from standard unipolar PAM signaling. This section now complements that analytical framework with a laboratory demonstration in which bipolar PAM4 is generated, transmitted over a short IM/DD optical path, and recovered by offline DSP. The objective is not to introduce new theory but to validate, under realistic transceiver bandwidth constraints and controlled OSNR conditions, that BPAM waveforms can be detected reliably and that modern data-driven equalization can be used as an effective nonlinear compensator in the oversampled direct-detection chain.

In particular, the experiment is organized around a fully instrumented, computer-controlled setup where OSNR is swept in a repeatable manner via optical noise loading, the received electrical waveform is captured at high bandwidth using a real-time oscilloscope, and a convolutional neural network (CNN) equalizer is applied offline to recover the transmitted sequence. The remainder of this section documents the measurement chain, the data pipeline, and the CNN equalizer used

¹⁴¹ X. Wu, A. Karar, K. Zhong, A. Lau, and C. Lu, “*Experimental demonstration of pre-electronic dispersion compensation in IM/DD systems using an iterative algorithm*”, *Optics Express*, vol. 29, no. 16, pp. 24735–24749, 2021.

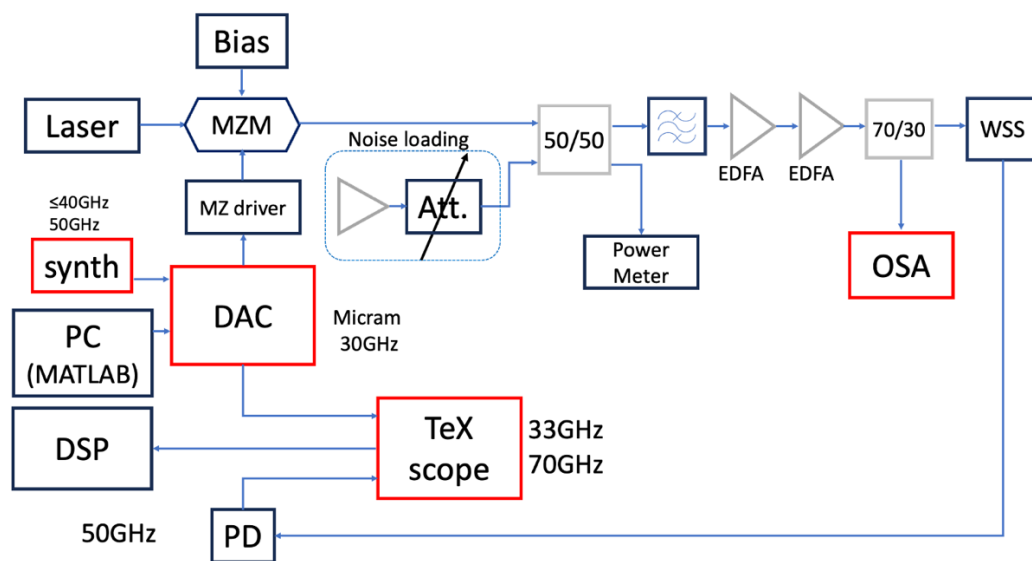
¹⁴² D. Che and X. Chen, “*Modulation Format and Digital Signal Processing for IM-DD Optics at Post-200G Era*”, *Journal of Lightwave Technology*, vol. 42, pp. 588–605, 2024.

in this work, together with the benchmark comparisons needed to interpret the results.

7.7.1 Experimental setup

Figure 7.x shows the complete laboratory setup used to experimentally evaluate bipolar PAM4 with direct detection and offline CNN equalization. The configuration follows a standard intensity-modulation/direct-detection (IM/DD) architecture, but it operates in a regime where the receiver intentionally maintains a fractionally spaced representation of the detected waveform (at least two samples per symbol after resampling). This allows the bipolar information embedded in the oversampled intensity to be recovered through DSP.

On the transmitter side, the digital waveform is generated offline on a laboratory PC using a MATLAB-based workflow and uploaded to a laboratory-grade DAC (MICRAM, 100 GSa/s, two-channel unit operated in single-channel mode for this experiment). The DAC output drives a broadband electrical amplification chain built with a Centellax three-stage broadband amplifier (model UA1L65VM), which functions as the Mach-Zehnder modulator (MZM) driver. The electrical drive signal is applied to an Exail MXAN-L40 intensity modulator (LiNbO₃ MZM). A stable DC bias is supplied to the MZM bias port to set the required operating point for the selected intensity-modulation format and to ensure repeatability during the OSNR sweep. The continuous-wave (CW) optical carrier is provided by a C-band tunable laser source (Yenista), set to an output power of 10 dBm at the desired wavelength.



To simulate controlled channel degradation without using long fiber spans, the optical path between transmitter and receiver consists of only a few meters of fiber patching. The main impairment is artificially introduced by optical noise loading. Specifically, amplified spontaneous emission (ASE) noise is generated optically and coupled into the signal path via a 50/50 optical coupler, as shown in Fig. 7.2. The noise-loading branch includes an optical amplifier stage and a programmable optical attenuator (JDS Uniphase HA series VOA) to set the injected noise level. After noise loading, the optical signal passes through optical amplification stages (EDFAs, as shown in Fig. 7.2) to achieve the desired received power level while maintaining a well-controlled OSNR sweep mechanism. A waveshaper (WSS) serves solely as a tunable optical filter in the link; in this experiment, it is not used for spectral shaping beyond filtering. For OSNR monitoring, a 70/30 coupler taps a fraction of the optical signal to an optical spectrum analyzer (Yokogawa AQ6389, 1200–1650 nm), while the remaining portion continues along the main path toward the receiver. Additionally, a power meter is connected at the intermediate tap point shown in Fig. 7.2 to monitor optical power during alignment and automated sweeps. At the receiver, the optically filtered signal is converted to the electrical domain using a u2t photodiode (model XPDV4120R). The photodiode's electrical output is directly connected to a Tektronix DPO77000-series real-time oscilloscope, operated in the appropriate bandwidth setting (33 GHz or 70 GHz, depending on

the specific measurement). This oscilloscope waveform serves as the primary data-capture point for the experiment. For each OSNR setting, the received time-domain trace is acquired and saved on the laboratory PC for later offline processing. No real-time DSP is performed in the measurement hardware; instead, synchronization, resampling, equalization (including the CNN equalizer), and BER evaluation are carried out offline in the DSP stage described in the following subsections. All instruments in the setup, including the DAC, VOA/noise-loading elements when addressable, the oscilloscope, and the OSA, are connected to the laboratory PC via LAN and/or GPIB, enabling centralized software control. Measurements are performed through an automated process where the OSNR reported by the OSA is adjusted in discrete steps (starting around 17 dB and increasing by approximately 1 dB each time). Multiple acquisition batches are collected at each OSNR level to assess repeatability and minimize the risk of outlier-driven conclusions.

7.7.2 Data acquisition and preprocessing pipeline (trace capture, OSNR sweep, storage, and normalization)

The experimental dataset used in this work is acquired directly from time-domain electrical waveforms captured at the photodiode output using a Tektronix DPO77000-series oscilloscope (Channel 1). For each operating point of the experiment, the optical path is first configured to the desired OSNR via the noise-loading branch, and the oscilloscope then records a sufficiently long waveform segment to support reliable offline processing. The entire measurement chain (including OSNR tuning and waveform capture) is controlled from the laboratory PC via LAN/GPIB connections, enabling repeatable acquisition across OSNR points.

The OSNR sweep is carried out in discrete steps. Starting from an OSNR of about 17 dB, the noise-loading level is adjusted to increase OSNR in approximately 1 dB increments. In practice, the OSNR reported by the OSA is not perfectly an integer due to measurement jitter and limited adjustment resolution; therefore, each dataset is tagged with the OSNR measured during acquisition. For each OSNR level, multiple acquisition batches are performed to assess repeatability and minimize susceptibility to occasional outliers from transient instrument conditions.

Each batch includes one oscilloscope capture saved to the laboratory PC for offline DSP processing.

To ensure that the acquired record is long enough for downstream processing and performance evaluation, the receiver script includes an explicit record-length consistency check. If the number of acquired samples is insufficient relative to the DAC waveform length and the oscilloscope sampling rate, the acquisition is rejected and the user is prompted to increase the oscilloscope record length. This prevents incomplete traces from entering the dataset and avoids subtle truncation artifacts during offline processing.

All captured traces are stored as MATLAB data files that contain the raw oscilloscope vector. The filename is automatically generated to include key measurement descriptors needed for later identification, such as the symbol rate, DAC sampling rate, oscilloscope sampling rate, BPAM4 amplitude configuration, optional pre-emphasis settings (if enabled), a timestamp, and the nominal OSNR tag used during the measurement session. This naming convention allows offline analysis to group measurements by OSNR and configuration without relying on external bookkeeping.

A final preprocessing step performed immediately after acquisition involves normalizing and quality-checking the alignment between the received waveform and the known transmitted reference used during waveform generation. In the receiver script, normalization is modeled as an affine calibration of the acquired waveform, meaning an offset removal and a scale adjustment applied to the captured trace before further analysis. When enabled, the calibration parameters are optimized numerically (using a simple unconstrained search) to maximize the estimated agreement/SNR between the normalized received trace and the corresponding reference sequence provided by the transmitter. The output of this step is a normalized received waveform that remains consistent across acquisition batches, along with a diagnostic overlay plot comparing a short segment of the normalized received signal to the reference. This process serves as an acquisition integrity check (confirming that the correct trace is loaded and scaling is consistent) and as a standardized input conditioning step for the subsequent DSP blocks described in the following subsections.

7.7.3 Mathematical foundations of the CNN equalizer

A convolutional neural network (CNN) equalizer can be interpreted as a data-driven, nonlinear generalization of a conventional tapped-delay-line equalizer. Instead of estimating a single linear filter, the CNN learns a hierarchy of local spatio-temporal feature extractors that operate directly on the received 1D waveform (or on short time windows of an oversampled waveform). In the context of IM/DD links, where bandwidth limitation, square-law detection, and component nonlinearities can produce waveform distortions that are not well captured by linear filtering alone, a 1D CNN provides a structured way to learn robust decision features from raw samples while preserving the locality that naturally matches finite-memory intersymbol interference (ISI). This 1D-CNN viewpoint, originally developed for generic time-domain signals, naturally generalizes to equalization because channel memory implies that each symbol decision depends primarily on a bounded neighborhood of samples rather than on the entire sequence¹⁴³.

Formally, let $x[n]$ denote the discrete-time received sample stream after acquisition and basic front-end conditioning (e.g., resampling to a fixed samples/symbol rate and amplitude normalization, as described in the previous subsection). A 1D convolutional layer computes a set of output feature maps by sliding learnable kernels over the input. For an input with C_{in} channels (feature maps) and an output with C_{out} channels, the pre-activation output of the m -th output channel can be written as

$$z_m[n] = b_m + \sum_{c=1}^{C_{\text{in}}} \sum_{i=0}^{K-1} w_{m,c}[i] x_c[n-i],$$

where K is the kernel length, $w_{m,c}[i]$ are the convolution weights, and b_m is a bias term. A pointwise nonlinearity $\phi(\cdot)$ then produces the layer output

¹⁴³ S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, “1D Convolutional Neural Networks and Applications: A Survey”, *Mechanical Systems and Signal Processing*, vol. 151, Art. no. 107398, 2021.

$$x'_m[n] = \phi(z_m[n]),$$

and subsequent layers repeat this operation on the resulting feature maps. In equalization term¹⁴⁴.

A central design quantity for CNN equalizers is the *receptive field*, i.e., the number of input samples that can influence a specific output sample or decision. If the CNN is used as a sliding-window equalizer (producing one output per time index), then the receptive field determines the maximum effective ISI memory that the model can exploit. For a stack of convolutional layers with stride s_ℓ , dilation d_ℓ , and kernel length K_ℓ at layer ℓ , the receptive field length R (in input-sample units) grows according to the effective kernel length $K_\ell^{\text{eff}} = d_\ell(K_\ell - 1) + 1$ and the cumulative downsampling introduced by strides. A common expression is

$$R = 1 + \sum_{l=1}^L (K_l^{\text{eff}} - 1) \prod_{j=1}^{\ell-1} s_j,$$

which reduces to $R = 1 + \sum_{\ell=1}^L (K_\ell - 1)$ when all strides are 1 and dilations are 1. In practice, for an equalizer, R should be selected to cover the dominant pre-cursor and post-cursor ISI span (in samples), while avoiding unnecessary growth that increases computation and may encourage overfitting. This “memory coverage” interpretation provides a direct bridge between classical equalizer design (choosing a tap length to cover channel memory) and CNN architecture design (choosing kernel sizes, number of layers, and dilations/strides to cover the same memory)¹⁴⁵.

Training the CNN equalizer follows standard supervised learning. A labeled dataset is formed by pairing received waveform segments (or sample windows) with their corresponding transmitted labels (bits or symbols). Denoting the dataset

¹⁴⁴ *Ibidem*.

¹⁴⁵ Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, “A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects”, IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 12, pp. 6999–7019, 2022.

by $\{(x^{(q)}, y^{(q)})\}_{q=1}^Q$ and the CNN output by $\hat{y}^{(q)} = f_{\theta}(x^{(q)})$, the parameters θ (all convolution weights and biases) are optimized by minimizing an empirical risk

$$\min_{\theta} \frac{1}{Q} \sum_{q=1}^Q \mathcal{L}(\hat{y}^{(q)}, y^{(q)}) .$$

For symbol/bit classification, a standard choice is cross-entropy loss; for waveform regression or soft-symbol estimation, mean-squared error (MSE) is often used. Gradients $\nabla_{\theta} \mathcal{L}$ are computed via backpropagation through the convolutional structure, and parameters are updated by stochastic gradient-based optimizers. In an equalization setting, this optimization can be interpreted as learning a nonlinear decision rule that best compensates ISI and distortion under the empirical channel statistics represented by the collected measurements, while the locality of convolution constrains the solution to be consistent with finite-memory signal formation¹⁴⁶.

7.7.4 CNN architecture used in this work (layers, kernel sizes, activation, normalization, output mapping)

The proposed equalizer is implemented as a compact 1D convolutional neural network (1D-CNN) that operates directly on the fractionally-spaced received waveform. After resampling the captured trace to 2 samples/symbol, each symbol decision is supported by a fixed-length temporal context window extracted from the received sequence. In the implementation used in this work, the window length is 20 samples, i.e., 10 symbols at 2 Sa/sym, and the network input for the k -th symbol is a real-valued vector $x_k \in \mathbb{R}^{20 \times 1}$ formed by concatenating consecutive $T/2$ -spaced samples starting at the symbol-aligned index.

Architecturally, the equalizer comprises three convolutional feature-extraction blocks, followed by a dense decision head. Each convolutional layer uses a kernel size of 3 and “same” padding, so the temporal length (20 samples) is preserved

¹⁴⁶ I. D. Mienye, T. G. Swart, G. Obaido, M. Jordan, and P. Ilono, “Deep Convolutional Neural Networks in Medical Image Analysis: A Review”, *Information*, vol. 16, no. 3, Art. no. 195, 2025.

through the convolutional stack while the number of channels increases with the number of learned filters. Specifically, the convolutional stack is:

$$\begin{aligned} &\text{Conv1D}(64, k = 3) \rightarrow \text{BN} \rightarrow \text{LeakyReLU}(\alpha = 0.1) \rightarrow \text{Dropout}(0.2) \\ &\text{Conv1D}(128, k = 3) \rightarrow \text{BN} \rightarrow \text{LeakyReLU}(\alpha = 0.1) \rightarrow \text{Dropout}(0.3) \\ &\text{Conv1D}(128, k = 3) \rightarrow \text{BN} \rightarrow \text{LeakyReLU}(\alpha = 0.1) \rightarrow \text{Dropout}(0.4) \end{aligned}$$

This structure is deliberately “filter-like”: the first layer learns a bank of short FIR-like templates on the raw input window, while deeper layers compose these local templates into higher-level features that capture short-memory ISI patterns and nonlinear distortions as they appear in the sampled direct-detection waveform. With three stacked $k=3$ convolutions (unit stride, no dilation), the *effective receptive field* spans 7 input samples, meaning each deep feature at a given time index is influenced by a neighborhood of $7 T/2$ -spaced samples; the overall decision is then formed by aggregating features across the full 20-sample context via the dense stage.

After the third convolutional block, the feature tensor (length 20, channels 128) is flattened and passed to a dense head:

$$\text{Flatten} \rightarrow \text{Dense}(256) \rightarrow \text{BN} \rightarrow \text{LeakyReLU}(\alpha = 0.1) \rightarrow \text{Dropout}(0.5)$$

The flattening step converts the learned spatio-temporal feature representation into a single feature vector suitable for symbol-level classification. Batch normalization is used both in the convolutional stack and in the dense head to stabilize training and reduce sensitivity to scale variations in the captured waveform, while dropout is included throughout to control overfitting given the strong correlation structure typical of oversampled IM/DD sequences.

The network output is a *two-dimensional sigmoid layer*, producing two independent probabilities:

$$\hat{p}_k = \sigma(Wh_k + b) \in (0,1)^2,$$

where h_k denotes the last hidden representation for the (k)-th input window. These two outputs correspond to (i) an amplitude-class bit and (ii) a sign/phase-related bit (as defined by the labeling used in the dataset). In the implementation, the amplitude labels are encoded as a binary target derived from a two-class amplitude index (mapped from $\{1,2\}$ to $\{0,1\}$). In contrast, the sign/phase labels are encoded as a binary target derived from $\{+1, -1\}$ (mapped to $\{0,1\}$). The use of sigmoid outputs reflects this two-bit, multi-label formulation; how these probabilities are converted into hard symbol decisions and then mapped back to BPAM-4 decisions is described in the inference/decision stage subsection (Sec. 7.7.6).

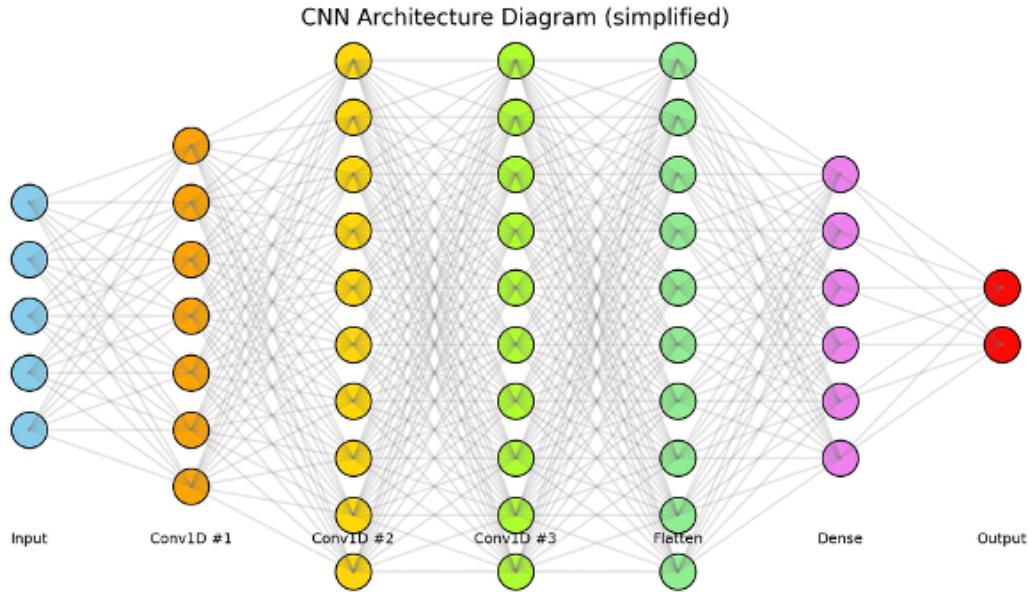


Figure 7.3 Simplified representation of the CNN architecture

7.7.5 Training procedure (loss, optimizer, regularization, early stopping, hyperparameters)

The CNN equalizer was trained using supervised learning with labeled examples taken from the experimentally collected oversampled waveform. After resampling the received sequence to a rate of two samples per symbol, the training set was built by sliding a fixed-length context window across the received samples. For symbol index k , the input vector is formed as

$$\mathbf{x}_k = [z_{rk}[2k], z_{rk}[2k + 1], \dots, z_{rk}[2k + W - 1]]^T, \quad W = 20,$$

so that each training example spans $W/2=10$ symbol intervals at the $T/2$ sampling grid. The corresponding supervision consists of two binary labels: an amplitude class and a phase/sign-related class. In our implementation, the amplitude label $A_k \in \{1,2\}$ is mapped to a binary target $a_k = A_k - 1 \in \{0,1\}$, while the phase label $B_k \in \{-1, +1\}$ is mapped to $b_k = (B_k + 1)/2 \in 0,1$. The network therefore learns a two-output binary classifier with target vector $\mathbf{t}_k = [a_k, b_k]$.

The model outputs two sigmoid activations $\hat{\mathbf{t}}_k = [\tilde{a}_k, \tilde{b}_k]$, interpretable as posterior probabilities for the two binary decisions. Training minimizes the binary cross-entropy over both outputs, averaged over the training set:

$$\begin{aligned} \mathcal{L} = -\frac{1}{N} \sum_{k=1}^N & (a_k \log(\hat{a}_k) + (1 - a_k) \log(1 - \tilde{a}_k) + b_k \log(\hat{b}_k) \\ & + (1 - b_k) \log(1 - \tilde{b}_k)) \end{aligned}$$

Optimization is performed with Adam using a learning rate of 10^{-4} and mini-batches of size 32. The dataset is split into 80% for training and 20% for testing/validation using a fixed random seed “random_state=42” to ensure repeatability of the split.

Regularization is implemented directly in the architecture through dropout and batch normalization. Dropout is applied after each convolutional block and after the dense layer, with rates 0.2, 0.3, 0.4, and 0.5, respectively. Batch normalization is used after each convolution and after the dense layer. To avoid overfitting and unnecessary training beyond convergence, early stopping is used with the validation loss as the monitored quantity; training is halted if the validation loss does not improve for 5 consecutive epochs (patience = 5), and the best-performing weights (minimum validation loss) are restored. The maximum number of epochs is set to 50, but in practice, early stopping typically terminates training earlier once the validation criterion saturates.

7.7.6 Inference and decision stage (soft/hard outputs, symbol decisions, BER evaluation)

At inference time, the received waveform is processed with the same windowing rule used during training. For each symbol index k , the input segment $\mathbf{x}_k$ of length $W=20$ samples is extracted starting at sample $2k$, reshaped to match the CNN input format $(W, 1)$, and passed through the trained network. The CNN produces two soft outputs $\hat{a}_k \in (0,1)$ and $\hat{b}_k \in (0,1)$ which represent the model's confidence for the amplitude bit and the phase bit, respectively. These sigmoid outputs can be used as soft information (e.g., for reliability-aware post-processing) because they provide a continuous score monotonically related to the likelihood of each class.

For the BER results reported in this work, hard decisions are formed by a fixed threshold at 0.5:

$$\tilde{a}_k = \begin{cases} 1, & \hat{a}_k \geq 0.5 \\ 0, & \hat{a}_k < 0.5 \end{cases}, \quad \tilde{b}_k = \begin{cases} 1, & \hat{b}_k \geq 0.5 \\ 0, & \hat{b}_k < 0.5 \end{cases}.$$

The decoded amplitude class is then recovered as $\hat{A}_k = \tilde{a}_k + 1 \in \{1,2\}$, while the decoded phase label is recovered as $\hat{B}_k = 2\tilde{b}_k - 1 \in \{-1, 1\}$. These two decisions jointly represent the estimated BPAM4 symbol label for time index k in the same labeling convention used to generate the training targets. When an explicit reconstructed multilevel symbol sequence is required, the amplitude class selects the corresponding magnitude level, and the phase decision assigns the associated bipolar sign according to the BPAM mapping used at the transmitter; the resulting $\hat{x}_k$ is then available for subsequent slicing or for producing waveform-level comparisons.

Performance is quantified by comparing the hard-decoded labels against the known transmitted labels associated with the acquired waveform segment. Let $\{a_k, b_k\}_{k=1}^N$ denote the ground-truth binary labels used for evaluation and $\{\tilde{a}_k, \tilde{b}_k\}_{k=1}^N$ their corresponding hard decisions. The total bit-error rate over the two output streams is computed as

$$\text{BER} = \frac{\sum_{k=1}^N (1[\tilde{a}_k \neq a_k] + 1[\tilde{b}_k \neq b_k])}{2N},$$

where $1[\cdot]$ is the indicator function. This definition reflects that each BPAM4 symbol carries two binary decisions in the adopted representation (amplitude and phase). In the OSNR sweep reported in the experimental results, the procedure is repeated independently at each OSNR point: the waveform is acquired, preprocessed, and windowed, inference is performed to produce $\tilde{a}_k, \tilde{b}_k$, and BER is computed by direct comparison against the corresponding transmitted label sequences for that capture.

7.7.7 Results and comparison to classical baselines

The performance of the proposed CNN equalizer was assessed using experimentally obtained waveforms and measured in terms of BER versus OSNR. The OSNR sweep was performed by controlled ASE loading and multiple acquisitions at each OSNR point; the resulting traces were processed offline with the same preprocessing chain and then equalized either by the proposed CNN or by a conventional adaptive linear equalizer. The comparison is therefore “like-for-like”: identical transmitted waveform generation, same optical/electrical front-end, same acquisition conditions, and identical post-acquisition processing, with the only difference being the equalization algorithm used in DSP.

The chosen baseline for benchmarking was an adaptive linear equalizer implemented as a fractionally spaced FFE operating on the oversampled received sequence, which has two samples per symbol. This selection is primarily driven by the BPAM direct-detection mechanism itself: oversampling results in two interleaved sample streams, with the main contributions coming from (i) symbol “self” terms and (ii) inter-symbol cross-terms. A two-branch linear FFE is a natural and widely used structure to reduce residual linear ISI and timing-phase sensitivity under such fractionally spaced sampling. In the benchmark implementation, the two FIR filters have different tap lengths (11 and 9 taps), are initialized as symmetric impulse responses, and are trained iteratively to minimize the discrepancy between received samples and known training symbols using a stochastic gradient descent

method. This baseline provides the strongest “classical” reference within the same oversampled receiver framework, without incorporating higher-order nonlinear kernels or machine learning components.

Baseline adaptive equalizer (two-branch fractionally spaced FFE).

As a benchmark, we compared the CNN equalizer with a conventional fractionally spaced feedforward equalizer (FFE) operating at two samples per symbol. The received oversampled sequence is split into two interleaved branches (even and odd phases), which are equalized by two independent real-valued FIR filters. Denoting the oversampled input by $x[m]$, we define the branch signals $x_{\text{even}}[n] = x[2n]$ and $x_{\text{odd}}[n] = x[2n + 1]$. The corresponding equalizer outputs are computed as

$$y_{\text{even}}[n] = \sum_{k=0}^{10} h_{\text{even}}[k] x_{\text{even}}[n - k], \quad y_{\text{odd}}[n] = \sum_{k=0}^8 h_{\text{odd}}[k] x_{\text{odd}}[n - k],$$

where the even branch uses 11 taps and the odd branch uses 9 taps, consistent with the two-phase T/2 sampling structure that is intrinsic to oversampled direct-detection processing.

Initialization and adaptive training.

Both FIR filters are initialized as symmetric impulse responses, i.e., a discrete-time delta located at the middle tap,

$$h[k] = \begin{cases} 1, & k = \lfloor L/2 \rfloor \\ 0, & \text{otherwise} \end{cases}$$

with $L=11$ for the even branch and $L=9$ for the odd branch. The coefficients are then updated adaptively using *stochastic gradient descent (SGD)* to minimize the error between the equalized output and the known transmitted training symbols. With a generic error signal $e[n]$ (computed from the mismatch between the equalizer output and the corresponding reference symbol at time (n)), the tap-update rule is

$$h[k] \leftarrow h[k] + \eta e[n]x[n - k],$$

where η is the adaptation step size. This update is applied independently to the even and odd branches using their respective inputs x_{even} and x_{odd} and tap vectors h_{even} and h_{odd} . This baseline, therefore, represents a strong classical solution for T/2-spaced IM/DD reception: it directly exploits the oversampled structure while remaining fundamentally linear, providing a principled benchmark against which the CNN's nonlinear feature-learning capability can be assessed.

Under these conditions, the CNN equalizer provided a clear performance improvement over the adaptive linear baseline. From the measured BER-versus-OSNR characteristics, the CNN achieves an OSNR advantage of approximately 5 dB at a BER of 10^{-2} relative to the adaptive linear equalizer. In other words, for the same target BER in this operating region, the CNN requires materially less OSNR, indicating that it is compensating a larger portion of the effective end-to-end impairment budget. This gain was observed despite the presence of non-negligible implementation loss attributed to practical laboratory limitations (finite bandwidth, component non-idealities, and residual distortions not fully removed by linear equalization).

In addition to BER, the CNN equalizer's internal decoding task, binary classification of amplitude and differential-phase (sign-relation) information used to reconstruct the bipolar PAM4 sequence, showed high stability across the tested datasets. The reported overall classification accuracy was 96.76% for joint amplitude/phase decoding, and this level was consistent across test sets, indicating that the trained model was not merely fitting a single capture but rather capturing repeatable structure in the received waveform distortions.

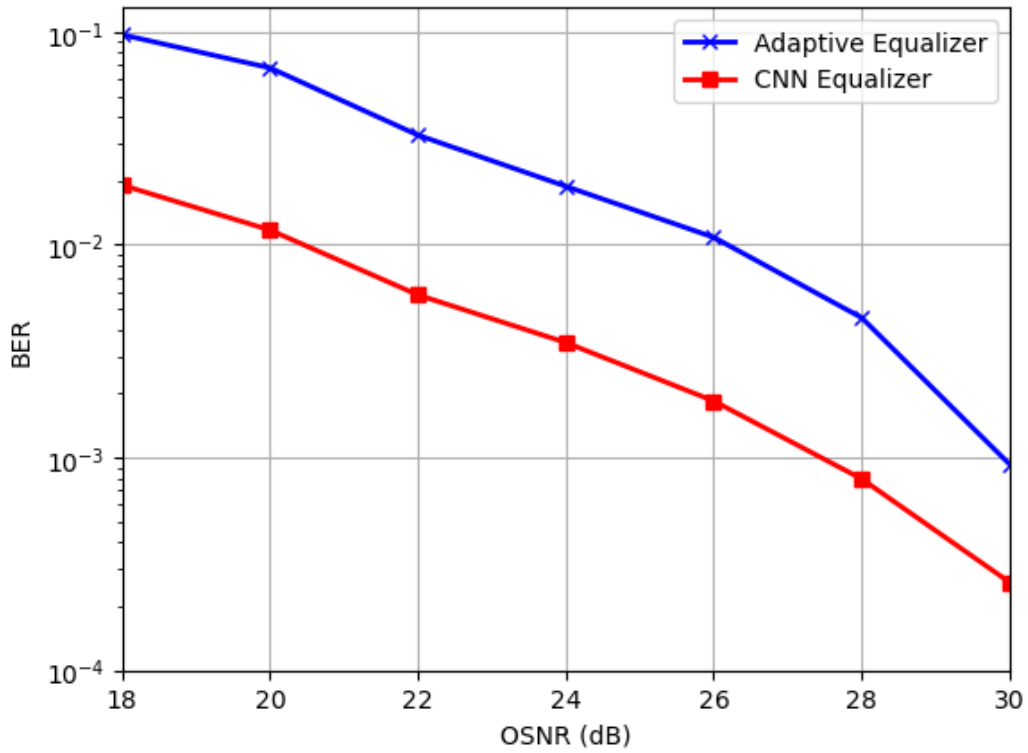


Figure 7.x Performance of the CNN model with respect to the adaptive equalizer

7.7.8 Practical discussion (complexity, latency, generalization, reproducibility)

From an implementation standpoint, the key practical distinction between the two equalizers is not only their impairment coverage (linear-only versus nonlinear-capable), but also how their computational cost, latency, and deployment workflow scale with symbol rate and channel memory. The classical benchmark used in this work is a *fractionally-spaced, two-branch FFE* operating on the interleaved even and odd samples at the $T/2$ grid. It employs 11 taps on the even branch and 9 taps on the odd branch, is initialized as a centered discrete-time impulse response, and is adapted through an SGD-type rule ($h[k] \leftarrow h[k] + \eta e[n]x[n - k]$). This structure has two practical strengths: (i) it is computationally light, dominated by a small number of multiply–accumulate operations per sample, and (ii) it is compatible with mature real-time implementations and can track slow drift through online adaptation. Its limitation is fundamental: it is a linear model that primarily mitigates linear ISI and timing-phase sensitivity. When the dominant penalties arise from composite nonlinear effects (e.g., modulator transfer characteristics, amplifier compression, square-law detection interactions, or mixed distortion mechanisms),

a short linear FFE may converge yet still leave a residual distortion floor, precisely the regime in which the CNN equalizer provided a measurable OSNR advantage in the experiment.

The CNN equalizer shifts the burden from online coefficient adaptation to offline (or semi-offline) training, followed by comparatively lightweight inference. In the demonstrated configuration, inference applies a small stack of 1D convolutional layers to a short context window (20 samples at the working oversampled rate) and maps the resulting features through a compact dense head to two binary outputs. Although the CNN contains more parameters than an 11/9-tap linear equalizer, its computations are highly regular (convolutions and pointwise nonlinearities) and therefore well-suited to acceleration on GPUs, and, importantly for future real-time feasibility, amenable to FPGA/ASIC mapping using streaming convolution engines. The practical trade-off is therefore not “linear versus ML” in the abstract, but rather whether the additional structured computation of the CNN is justified by a tangible performance gain in the impairment regime of interest; in this study, the observed OSNR improvement indicates that the added modeling capacity translates into a meaningful link-budget benefit.

Latency in a streaming receiver is governed by the equalizer memory and by whether decisions rely on future samples. The two-branch fractionally-spaced FFE is naturally causal when implemented as a conventional FIR filterbank; its decision latency is essentially the group delay implied by the selected tap spans. The CNN equalizer, instead, operates on a finite-length window. In a real-time deployment, this implies buffering at least the window length before producing a decision. If the window is defined symmetrically around the decision instant, the method introduces a look-ahead delay (non-causality); if the window is defined using only current and past samples, the model can be made causal at the cost of either reduced performance or retraining under a causal window definition. In the present work, equalization is performed offline on captured traces, so buffering does not constrain the reported results; nevertheless, windowing and causality become first-order design parameters when targeting hardware implementation at very high symbol rates.

Generalization is another practical dimension in which ML equalizers differ from classical adaptive filters. A linear adaptive equalizer can track moderate channel variations continuously as long as its linear model remains valid. A CNN, in contrast, generalizes well within the distribution represented in its training data, but may degrade under distribution shift (different component responses, different operating point, different noise statistics, or changes in bandwidth constraints). Robust deployment, therefore, typically benefits from training across diverse operating conditions (e.g., multiple OSNR points and representative impairment variations), and, in scenarios where retraining is feasible, from periodic updates or transfer-learning strategies that adjust only a subset of layers. In this study, the high and consistent decoding accuracy across repeated acquisitions within the same laboratory configuration indicates stable generalization under the tested conditions; it does not, by itself, guarantee portability across substantially different hardware chains without recalibration.

Finally, reproducibility in experimental ML-based equalization depends on both measurement determinism and documentation completeness. On the positive side, the experiment used automated OSNR stepping and automated waveform capture, reducing human-induced variability and enabling repeated measurements at each operating point. The baseline equalizer is also specified at the structural and algorithmic level (two-branch T/2 FFE, 11/9 taps, impulse initialization, SGD-type adaptation). For full external reproducibility, however, a third party would still require access to representative captured datasets (or a defined data-release policy), explicit preprocessing definitions (resampling method, normalization, windowing rule, and labeling), and the training configuration and artifacts for the CNN (random seeds, optimizer settings, early-stopping criterion, and the final trained weights). Where raw datasets or control scripts cannot be shared, the most practical alternative is to publish a minimal runnable pipeline operating on an anonymized subset of traces, together with sufficient metadata to confirm that the same signal model and label conventions are used.

A key opportunity to further improve both performance and reproducibility is to integrate the developed *digital surrogate* into the equalizer-design workflow. Because the surrogate provides a modular, physically grounded emulation of the

end-to-end IM/DD chain, including bandwidth limitations, component nonlinearities, square-law detection effects, and controllable noise/impairment injection, it can be used to generate large, labeled datasets under systematically varied conditions that are impractical to explore exhaustively in the laboratory. This enables (i) principled data augmentation and domain-randomization to harden the CNN against distribution shift, (ii) controlled ablation studies to identify which impairment families drive the CNN's advantage over linear equalization, (iii) architecture and hyperparameter sweeps (window length, receptive field, regularization) guided by realistic channel statistics, and (iv) calibration loops in which surrogate parameters are fitted to a small set of measured traces, after which the surrogate can be used to extrapolate performance across OSNR, bandwidth, or operating-point ranges. In this sense, the surrogate acts not only as a simulator, but as a repeatable experimental proxy that can reduce lab time, improve the statistical robustness of reported results, and accelerate the co-design of modulation, receiver DSP, and ML equalization for bipolar PAM with direct detection.

Chapter 8

Conclusions and Perspectives

This chapter summarizes the thesis results and contextualizes them within broader mixed-signal short-reach IM-DD communications and learning-aided optimization. It begins by reviewing the contributions and key technical findings, then responds to the research questions from Chapter 1, demonstrating how each thesis objective has been achieved. Finally, it discusses current limitations impacting the fidelity and scope of the digital surrogate, and proposes future directions for more accurate experimental calibration, broader system coverage, and integrated optimization workflows.

8.1 Executive summary of the thesis

This thesis addresses the practical challenge that modern short-reach IM-DD optical links operate in a regime in which performance is influenced by strongly coupled impairments across the transmitter, optical path, and receiver. Instead of treating these subsystems as separate and optimizing them with partially decoupled models, the thesis contends that realistic optimization requires an end-to-end approach where mixed-signal non-idealities are explicitly represented, parameterized, and observable at multiple internal points.

To this end, the thesis introduces and implements a *digital surrogate*: a modular, device-aware, end-to-end framework for IM-DD PAM and bipolar-PAM variants, designed not only as a simulator but as a structured generator of analysis artifacts and ML-ready datasets. The surrogate formalizes disciplined interfaces, parameter governance, time and sampling consistency, metadata separation for visualization and learning, and a validation strategy that is compatible with hardware calibration and experimental benchmarking.

The core messages of the thesis can be summarized as follows:

- **End-to-end coupling as a first-order design constraint.**

The thesis shows that, at high symbol rates and limited bandwidth margins, IM-DD links are best viewed as a finite-memory (often nonlinear) discrete-time channel whose behavior results from the interaction of DAC/driver/modulator response, dispersion, noise loading, and receiver front-end plus ADC sampling and quantization. The proposed digital surrogate clarifies these couplings, making them open to systematic analysis and optimization instead of ad-hoc tuning.

- **A modular architecture that preserves physical meaning while enabling reproducibility.**

The surrogate is designed as a block-modular chain with standardized I/O contracts, clear domain labeling, and strict parameterization, avoiding hidden assumptions and hard-coded constants. This approach allows for controlled ablation studies, repeatable sweeps across operating points, and reliable comparisons of different modeling and DSP options, while remaining adaptable to additional devices and formats.

- **Device-aware impairment modeling is expressed as controllable “knobs.”**

Dominant non-idealities, especially in the mixed-signal transmitter and receiver chains, are modeled as parameterized blocks with clear physical meaning (e.g., quantization standards, INL/DNL, reconstruction effects, bandwidth limits, noise densities, sampling jitter). This approach enables optimization and sensitivity analysis to be connected to hardware-relevant settings. The structure supports both forward simulation and calibration-based refinement.

- **ML readiness through structured metadata and observation points.**

The surrogate is designed to generate not only waveforms but also context, including operating-point descriptors, impairment parameters, observation tags, and derived metrics that support learning-assisted equalization and optimization with improved interpretability and control. The metadata strategy is explicitly aligned with reproducible dataset creation and aims to

reduce domain shift by making sources of mismatch measurable rather than implicit.

- **A demonstration that connects modeling, DSP, and experiment.**

The thesis features a receiver-side DSP treatment and an experimental study that together demonstrate how the surrogate perspective connects theory and practice. Specifically, the bipolar-PAM case study shows how oversampling-enabled direct detection changes the recoverable information and provides a clear context in which ML equalization can be compared to a traditional baseline under controlled OSNR sweeps and realistic data acquisition constraints.

8.2 Answers to the research questions

8.2.1 RQ1: How can a modular end-to-end surrogate preserve physically meaningful coupling across Tx–channel–Rx?

This thesis demonstrates that maintaining end-to-end coupling within a modular framework involves more than just assembling independent component models; it requires a rigorous approach to interfaces and shared invariants. The proposed digital surrogate accomplishes this by enforcing a standardized block contract, explicit domain transitions (digital/electrical/optical), a canonical waveform container, and a single authoritative time base. These design choices ensure impairments are not implicitly reinterpreted at each block boundary and that sampling-rate transformations, filtering, and nonlinear mappings stay consistent throughout the chain. In this way, the surrogate preserves the causal pathway by which transmitter mixed-signal behavior, optical propagation/noise loading, and receiver sampling/quantization collectively shape the discrete-time channel observed by DSP.

This answer supports the thesis premise that meaningful optimization cannot be reliably performed on partially decoupled models. By construction, the surrogate enables systematic exploration of coupled regimes and therefore provides a stable foundation for both DSP evaluation and learning-assisted optimization across operating points.

8.2.2 RQ2: How can dominant non-idealities be modeled at sample-level fidelity while remaining feasible and calibratable?

The thesis shows that feasibility and calibratability are achieved by combining (i) a fidelity ladder that prioritizes impairments with the strongest end-to-end impact, (ii) parameterized models with clear, measurement-aligned interpretation, and (iii) execution modes that allow for lighter computational runs when appropriate. The surrogate focuses on sample-level representations where they are needed to capture memory and nonlinear coupling (e.g., reconstruction effects, bandwidth-limited responses, SOA dynamics, receiver front-end filtering, sampling perturbations), while keeping each impairment modular and switchable. By expressing device behavior through physically meaningful parameters (rather than opaque black-box transformations), the framework stays compatible with calibration and systematic refinement as new measurements become available.

The resulting surrogate is sufficiently detailed to reproduce coupled impairment behavior relevant to receiver DSP and optimization, while remaining extensible and calibratable, two requirements that are commonly in tension in high-speed communication modeling.

8.2.3 RQ3: How can the surrogate generate ML-ready datasets and metadata for controllable and interpretable optimization?

This thesis considers ML readiness as an integral part of the surrogate rather than an afterthought. Specifically, the surrogate is designed to generate waveforms along with structured metadata that encodes operating conditions, impairment configurations, observation locations, and performance metrics. This setup supports dataset creation where labels and features are linked to physical controls, enabling learning models to be trained and tested in ways that maintain interpretability and allow for controlled assessments of generalization. Separating visualization metadata from learning and optimization metadata further promotes clean experimental design: the surrogate can be used for exploratory analysis, systematic dataset generation, or optimized execution without mixing these purposes. By design, the surrogate facilitates hybrid workflows in which model-based understanding and data-driven learning strengthen each other, reducing reliance on

fragile black-box approaches and clarifying the relationship between performance improvements and physical/system parameters.

8.2.4 RQ4: How should surrogate fidelity be validated against experiments, and how can mismatch/uncertainty be managed?

The thesis employs a validation approach that considers a mismatch as a valuable signal rather than a flaw to conceal. Internally, the surrogate uses consistency checks and regression-style “golden cases” to verify that updates do not silently break invariants throughout the chain. Externally, the thesis demonstrates an experimental workflow where data acquisition, OSNR-controlled impairment loading, and preprocessing are thoroughly documented, enabling meaningful comparisons between model predictions and actual measurements. Although a comprehensive uncertainty quantification framework remains an open research area, the surrogate’s modular design helps identify mismatch sources, facilitating incremental calibration and targeted refinement of the most critical components. This approach offers a practical pathway from simulation to experimentation: the surrogate can adapt through measurement-informed updates while preserving reproducibility and traceability, thus reducing domain shift risks in learning-assisted optimization.

8.3 Achievement of the thesis objectives

8.3.1 O1: Architecture: define a rigorous, modular digital-surrogate architecture

The thesis presents a structured architecture for an end-to-end IM-DD digital surrogate that emphasizes modularity, reproducibility, and physical interpretability. Key to this achievement is the definition of a standardized block interface contract, the enforcement of explicit domain transitions, the use of a canonical waveform container, and the management of global time and sampling consistency. The architecture also introduces a layered parameter hierarchy that removes hard-coded assumptions and allows systematic sweeps, along with a separation of visualization metadata from learning and optimization metadata to avoid role conflation.

The architecture is shown to support consistent end-to-end simulation, intermediate observation points, lightweight execution modes, and validation hooks that preserve correctness under iterative development.

8.3.2 O2: Device-aware modeling: implement dominant non-idealities across Tx, channel, and Rx

The thesis implements a prototype chain in which dominant mixed-signal and electro-optic non-idealities are modeled as explicit, parameterized blocks. On the transmitter side, this includes DAC-related impairments and reconstruction effects, driver limitations, and modulator nonlinearity. The optical path includes propagation and noise-loading mechanisms aligned with IM-DD practice. On the receiver side, the front-end response and ADC-chain effects are modeled with emphasis on bandwidth limitation, noise contributions, sampling considerations, and quantization behavior. This representation is organized to preserve end-to-end coupling and to keep device behavior visible as controllable knobs.

The chain reproduces a finite-memory discrete-time channel at the DSP interface whose properties emerge from the interaction of transmitter, optical, and receiver impairments rather than from isolated abstractions.

8.3.3 O3: ML-readiness, enable structured dataset generation and interpretable optimization

The thesis defines the surrogate not only as a simulator but also as a structured data generator. The framework enables systematic creation of waveforms with associated metadata describing operating points, impairment configurations, observation locations, and performance metrics. This design enhances learning-assisted processing by improving interpretability through linking training data to physical and system-level parameters. The separation of metadata roles further allows the same surrogate to support exploratory visualization, controlled dataset generation, and optimization experiments without causing ambiguity about what is measured and stored.

The ML case study demonstrates that fractionally spaced observations and structured processing pipelines enable evaluation of learning-assisted equalization under controlled experimental conditions and comparison with classical baselines.

8.3.4 O4: Validation and demonstration: connect the surrogate perspective to the receiver DSP and experiment

The thesis presents a receiver-side DSP treatment aligned with the surrogate perspective and showcases an experimental pipeline rooted in realistic acquisition constraints. The bipolar-PAM study emphasizes how oversampling-enabled direct detection changes the structure of recoverable information and drives the need for specialized receiver processing. The experimental methodology, including OSNR sweeps, trace management, and preprocessing, ensures reproducible evaluation, while the ML equalizer case study offers a clear example of learning-assisted processing within the broader surrogate framework.

A measurable performance comparison between an ML-based equalizer and a classical baseline under controlled impairment loading, with explicit documentation of the system chain and data handling.

8.4 Consolidated list of contributions

C1: A modular, end-to-end digital surrogate for short-reach IM-DD systems.

This thesis presents an end-to-end digital surrogate framework that maintains coupled transmitter–channel–receiver effects while ensuring strict modularity and parameter control. The framework aims to facilitate reproducible experiments and systematic exploration of operating points, providing clear visibility of where and how impairments are introduced and monitored.

C2: A disciplined runtime structure enabling consistency of time, sampling, and interfaces.

The thesis defines a clear process for mixed-signal simulation using a standard waveform format, a single consistent time base, and uniform block agreements. This process helps prevent unnoticed errors when blocks are enabled, disabled, or replaced, and it supports internal validation checks and regression workflows.**C3:**

Device-aware impairment modeling with controllable, calibratable parameters.

The thesis develops device-aware transmitter and receiver prototypes where key non-idealities are described as physically meaningful parameters instead of opaque transformations. This approach supports both sensitivity analyses and calibration-based refinement. The modeling choices focus on impairments that most affect performance in bandwidth-limited, short-reach IM-DD links and that show strong end-to-end coupling.

C4: An ML-ready data and metadata strategy for interpretable optimization.

The thesis considers ML readiness as a design goal for the surrogate, allowing for structured dataset creation where waveforms are paired with metadata describing operating points, impairment states, and observation sites. This facilitates learning-assisted equalization and optimization workflows that stay interpretable and controllable through physical knobs.

C5: A receiver DSP and experimental case study for bipolar-PAM with learning-assisted equalization.

The thesis offers an experimental demonstration and receiver-side analysis, showing how oversampling-enabled direct detection supports bipolar information recovery and encourages learning-assisted equalization. It establishes a practical, reproducible evaluation pipeline under OSNR-controlled impairment conditions and compares machine learning and traditional methods within a consistent signal-processing framework.

8.5 Limitations

The thesis presents a validated prototype of an end-to-end digital surrogate and shows its usefulness for DSP and ML evaluation; however, several limitations restrict the current scope and inspire future research.

Scope limitation (formats and link classes).

The surrogate is developed with emphasis on short-reach IM-DD PAM and bipolar-PAM variants and therefore does not yet claim generality across all modulation formats and receiver architectures. Extending to broader classes, such

as coherent links or multi-carrier variants, requires additional modeling layers and validation strategies, even if the architectural discipline remains applicable.

Calibration breadth and hardware variability.

While the surrogate is designed to be calibratable, the thesis demonstration relies on a specific experimental setup and does not yet account for multi-device variability, temperature dependence, aging, or manufacturing spread comprehensively. Addressing these factors requires multi-instrument campaigns and statistically grounded parameter identification routines.

Computational and workflow constraints.

Sample-level modeling of coupled impairments can be computationally intensive, especially when extensive parameter sweeps or large dataset creation is needed. Although the framework supports modular enable/disable patterns and lightweight execution modes, more work is required to formalize acceleration techniques (e.g., model-order reduction, surrogate-within-surrogate approximations, parallel dataset generation).

Uncertainty quantification and domain-shift mitigation.

The thesis considers mismatch as observable and localizable, but a comprehensive uncertainty quantification framework that can propagate parameter uncertainty to performance distributions and support robust optimization remains an open area of research. This framework is especially important when ML models are trained on synthetic data and used on measured data, where domain shift is unavoidable.

8.6 Perspectives and future work

The digital surrogate developed in this thesis is intentionally designed as a *foundation* rather than a closed artifact: it establishes modularity, parameter governance, and observation discipline in a form that can be progressively calibrated, expanded, and integrated with measurement workflows. Accordingly, the most meaningful future work is not a single extension. Still, a staged transition in capability, moving from a calibrated simulator to a *digital shadow* (measurement-aligned and continuously updated), and ultimately toward a *digital*

twin of the IM-DD system in which models and measurements co-evolve under closed-loop synchronization and validation. The roadmap below reflects this progression and prioritizes the research directions that most directly enhance fidelity, transferability, and relevance to optimization.

8.6.1 Near-term: full calibration and upgrade to a digital shadow

The immediate goal is to fully calibrate the current framework using laboratory measurements and then upgrade it to a digital shadow, meaning a model that not only captures the entire process but is also systematically aligned with observed behavior and updated as new data becomes available. This step is crucial for reducing domain shift, enabling accurate parameter identification, and turning the surrogate into a dependable tool for design exploration and optimization. Specifically, the near-term efforts focus on three closely related actions.

First, the impairment library must be completed across all components, with clear separation of mechanisms that are currently combined or simplified. Specifically, the model should include all significant nonlinearities relevant to the operating regimes of interest. For mixed-signal converters, this involves implementing separate, physically grounded noise sources for DAC and ADC, rather than relying on coarse lumped abstractions, and ensuring their placement along the chain matches the actual signal path and bandwidth shaping. This refinement is essential because converter impairments often interact with bandwidth limitations and nonlinear transfer functions, influencing the effective channel memory as seen by the DSP.

Second, the electro-optic modulation stage should be expanded beyond a single baseline model. While a standard MZM transfer captures key memoryless nonlinearity, practical systems increasingly rely on more complex electro-optic modulators and drive architectures. Future work should include multiple MZM model classes, ranging from advanced lumped models to digital surrogates of segmented modulators that represent segment-level drive, distributed effects, and programmable shaping behavior. This direction is strategically crucial because segmented modulators serve as a natural bridge between hardware adjustments and

waveform-level optimization, and their surrogate models enable design-space exploration that is otherwise costly in experimental settings.

Third, the SOA and channel models should include all relevant nonlinear regimes, including those that cause pattern dependence and nonlinear memory. For SOAs, this means expanding reduced-order models to capture more complex dynamics as needed by measurements, and ensuring that ASE modeling aligns with filtering and detection so that beat-noise and signal–noise interactions are accurately represented. For the channel, going beyond baseline dispersion and attenuation into regimes relevant to the intended lab scenarios guarantees that calibration does not “compensate for missing physics” by adjusting unrelated parameters.

Collectively, these steps transform the surrogate into a digital shadow: a calibrated representation that supports trustworthy performance prediction under controlled parameter changes and provides a stable base for subsequent optimization studies.

8.6.2 Mid-term: high-fidelity multi-path extensions and polarization-aware modeling

The mid-term goal is to expand the surrogate from a single-lane end-to-end chain to a polarization-aware, multi-path model with accuracy similar to laboratory measurements. The most important focus is to model dual polarization and systems with multiple electro-optic modulators (EOMs), not as idealized parallel branches, but with enough realism to reproduce the coupling mechanisms seen in practice. This involves careful alignment of impulse responses, bandwidth limits, phase and bias behaviors, amplitude/phase imbalances, and cross-coupling effects that are observable in measurements when the system operates near its performance limits.

A key requirement of this horizon is measurement-driven fidelity: the goal is not just to add complexity, but to accurately reproduce the main features of lab data with traceable parameter interpretations. This means (i) expanding the calibration process to multi-channel identification, (ii) developing consistency constraints across polarization branches and multiple EOM stages, and (iii) ensuring that model extensions enhance predictive capability rather than simply adding compensating

degrees of freedom. Achieving this level of realism would make the surrogate significantly more useful for evaluating advanced DSP and learning-based strategies under conditions that closely resemble experimental data deployments.

8.7.3 Long-term: upgrade to a full digital twin and extension toward coherent and cross-domain surrogates

The long-term goal is to develop the framework into a complete digital twin of an IM-DD optical communication system. In this phase, the surrogate transitions from a calibrated shadow to a synchronized counterpart capable of supporting closed-loop updates, systematic uncertainty management, and ongoing validation as devices drift, are reconfigured, or replaced. Methodologically, this involves linking the surrogate with automated measurement pipelines, establishing robust parameter identification loops, and integrating uncertainty-aware modeling so that predictions and optimization decisions include quantified confidence.

Furthermore, the long-term trajectory naturally extends beyond IM-DD in two directions. First, the framework should be generalized to coherent digital surrogates using device-based models, maintaining the same architectural principles while expanding the physical scope to include coherent transceivers, their associated impairments, and DSP chains. Second, the surrogate philosophy can be applied to wireless communication setups that reuse standard devices and measurement infrastructure, enabling cross-domain digital surrogates where shared components (such as mixed-signal converters, RF front-ends, and signal-processing toolchains) are modeled consistently across optical and wireless contexts. This approach aligns strategically with the broader goal of creating reusable, device-aware digital twins that support optimization and validation across diverse communication systems.

Taken together, these future directions articulate a coherent progression: from a modular end-to-end digital surrogate, to a calibrated digital shadow, and ultimately to a full digital twin capable of supporting closed-loop optimization and deployment-relevant decision-making in both optical and, potentially, cross-domain communication infrastructures.

Bibliography

Achtenberg, K., and Bielecki, Z., “Comparative Analysis of the Selected Photoreceiver Input Stages in Terms of Noise”, *Sensors*, 25(5), 1359 2025.

Agrawal, G. P., “Fiber-Optic Communication Systems”, Wiley, 5th ed., Wiley, 2021.

Ahmed, H., and Lofstead, J., “Managing Randomness to Enable Reproducible Machine Learning,” in *Proceedings of the 5th International Workshop on Practical Reproducible Evaluation of Computer Systems*, pp. 15-20, 2022.

Ahmed, A. H., Moznine, A., Lim, D., Kim, Y. G., Rylyakov, A., and Shekhar, S., “A Dual-Polarization Silicon-Photonic Coherent Transmitter Supporting 552 Gb/s/Wavelength”, *IEEE Journal of Solid-State Circuits*, vol. 55, no. 9, pp. 2597-2608, 2020.

Ait Ouallane, A., et al. “Fusion of engineering insights and emerging trends in traffic management systems.” *Information Fusion*, vol 88, 218-248, 2022.

Akrout, M., Feriani, A., Bellili, F., Mezghani, A., & Hossain, E., “Domain Generalization in Machine Learning Models for Wireless Communications: Concepts, State-of-the-Art, and Open Issues,” *IEEE Communications Surveys & Tutorials*, vol. 25, pp. 3014–3037, 2023.

Alanazi, F., & Alenezi, M. “A framework for integrating intelligent transportation systems with smart city infrastructure.” *Journal of Infrastructure, Policy and Development*, 8(5), 3558, 2024.

Almaviva S.p.A. “MOOVA: Almaviva’s platform for sustainable, multimodal, as-a-service mobility,” *TTS Italia Newsletter*, 28 June 2022.

Almaviva S.p.A. “Play it safe with MOOVA Passenger Flow,” *TTS Italia Newsletter*, 28 April 2022.

Almaviva S.p.A. “The Development of Sustainable Transportation in the Cities of the Future - Almaviva with Roma Capitale at the Smart City Expo World Congress in Barcelona,” *Company press release*, 6 November 2024.

Almaviva S.p.A. "Citizenship & Government - Urban Future," Corporate web page, accessed 2025.

Arendt, P. D., Apley, D. W., and Chen, W., "Quantification of model uncertainty: Calibration, model discrepancy, and identifiability", *Journal of Mechanical Design*, vol. 134, no. 10, 100908, 2012.

Atieh, A., and Raytchev, M., "Optical communication system (OptiSystem) software enabling remote education and teaching," *Proceedings of SPIE, Seventeenth Conference on Education and Training in Optics and Photonics (ETOP 2023)*, 2023.

Babar, M., and Roberts, G., "A Multi-Stage RC Compensation Technique for Decoupling the Transimpedance and BW: Creating High Speed and Low Noise TIA Designs", *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 72, pp. 3847–3860, 2025.

Bahrami, H., Sepehrian, H., Park, C., Rusch, L., & Shi, W., "Time-Domain Large-Signal Modeling of Traveling-Wave Modulators on SOI", *Journal of Lightwave Technology*, 34(11), 2812-2823, 2016.

Balestrieri, E., "DAC Time-Domain Specifications Toward St...surement", *IEEE Transactions on Instrumentation and Measurement*, vol. 57, no. 7, pp. 1290-1297, Jul. 2008.

Barricelli, B. R., Casiraghi, E., and Fogli, D., "A survey on digital twin: Definitions, characteristics, applications, and design implications", *IEEE Access*, vol. 7, pp. 167653-167671, 2019.

Belykh, V. V., Rautert, J., Mikhlin, V. S., and Gubenko, A. E., "Bit Error Rate in WDM Data Transmission Links With Semiconductor Optical Amplifier", *Journal of Lightwave Technology*, vol. 43, pp. 1076–1088, 2025.

Beshr, A., and Aly, M., "Characterization of wideband semiconductor optical amplifier: numerical analysis and simulation", *Optical and Quantum Electronics*, vol. 55, 287, 2023.

Borraccini, G., Huang, Y., D'Amico, A., De Lima, T., Ip, E., Curri, V., Wang, T., and Asahi, K., "Optical Line Physical Parameters Calibration in Presence of EDFA Total Power

Monitors”, 2024 Optical Fiber Communications Conference and Exhibition (OFC), paper M3I.5, 2024.

Breque, K., De Nul, L., Petridis, A., “Industry 5.0 - Towards a sustainable, human-centric and resilient European industry,” European Commission, Brussels 2021.

Brodny, J. et al, “Assessing Human-Centric, Resilient, Sustainable and Smart Transformation in the EU-27 Countries: A New Balanced Composite Index Method”, *Journal of Open Innovation: Technology, Market, and Complexity*, 100685, 2025.

Buchwald, A., “High-speed time interleaved ADCs”, *IEEE Communications Magazine*, vol. 54, pp. 71–77, 2016.

Bıyık, C., Abareshi, A., Paz, A., Ruiz, R. A., Battarra, R., Rogers, C. D. F., & Lizarraga, C. “Smart mobility adoption: A review of the literature.” *Journal of Open Innovation: Technology, Market, and Complexity*, 7(2), 146, 2021.

Caballero-Gaudes, C., Reynolds, R., “Methods for cleaning the BOLD fMRI signal”, *NeuroImage*, vol. 154, pp. 128–149, 2017.

Cariolaro, G., Corvaja, R., and Pierobon, G., “Exact performance evaluation of lightwave systems with optical preamplifier”, *European Transactions on Telecommunications*, vol. 5, pp. 757–766, 1994.

Chakrabarty, A., Maddalena, E., Qiao, H., and Laughman, C., “Scalable Bayesian optimization for model calibration: Case study on coupled building and HVAC dynamics”, *Energy and Buildings*, vol. 240, 111460, 2021.

Che, D., and Chen, X., “Modulation Format and Digital Signal Processing for IM-DD Optics at Post-200G Era”, *Journal of Lightwave Technology*, vol. 42, pp. 588–605, 2024.

Chen, L., Krongold, B., and Evans, J., “Performance Analysis for Optical OFDM Transmission in Short-Range IM/DD Systems”, *Journal of Lightwave Technology*, vol. 30, no. 7, pp. 974-983, 2012.

Chen, Z., Dai, X., Nie, J., Zhang, S., Zhou, J., Zhang, J., Qiu, Y., Luo, M., Yang, Q., Deng, L., Cheng, M., Qiu, K., and Liu, D., “Complexity curtailed MLSE equalizer based on a

compact look-up-table for C-band DSB IM/DD transmission,” *Optics Letters*, 47(21), 5606–5609, 2022.

Chen, Z., Nie, J., Zhang, S., Yang, Q., Dai, X., Deng, L., Cheng, M., Liu, D., “56-Gb/s/ λ C-band DSB IM/DD PAM-4 40-km SSMF transmission employing a multiplier-free MLSE equalizer,” *Optics Express*, vol. 30(7), pp. 11275-11287, 2022.

Cheon, S., Kweon, S., Aberra, A., Je, M., and Ha, S. “Optimal Parameter Design of DAC-Based Sinusoidal Signal Generators for Electrical Impedance Spectroscopy”, *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1-11, 2023.

Chou, F., and Hung, C., “Glitch Energy Reduction and SFDR Enhancement Techniques for Low-Power Binary-Weighted Current-Steering DAC.” *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 24, pp. 2407-2411, 2016.

Città metropolitana di Roma Capitale. “Piano Urbano della Mobilità Sostenibile (PUMS) della Città metropolitana di Roma Capitale,” Official PUMS documentation and portal, 2023-2024.

Cooke, M., and Nicolici, N., “An Embedded Architecture for DDR5 DFE Calibration Based on Channel Stimulus Inversion,” *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 33, 793–806, 2025.

Cui, W., Huo, J., Zhu, J., Tao, J., Qin, P., Zhang, X., Bai, H., “Machine Learning-Assisted Mitigation of Optical Multipath Interference in PAM4 IM-DD Transmission Systems,” *Photonics*, 12(4), 310, 2025.

Da Dalt, N., Harteneck, M., Sandner, C., and Wiesbauer, A., “On the jitter requirements of the sampling clock for analog-to-digital converters,” *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 49, no. 9, pp. 1354-1360, Sept. 2002.

Da Ros, F., Ranzini, S., Bülow, H., & Zibar, D., “Reservoir-Computing Based Equalization With Optical Pre-Processing for Short-Reach Optical Transmission,” *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 26, pp. 1–12, 2020.

Daponte, P., De Vito, L., Iadarola, G., and Rapuano, S., “A reduced-code method for Integral Nonlinearity testing in DACs”, *Measurement*, vol. 182, art. 109764, 2021.

De Melo, A., and Petermann, K., “On the amplified spontaneous emission noise modeling of semiconductor optical amplifiers”, *Optics Communications*, vol. 281, pp. 4598–4605, 2008.

Declercq, J., Niu, S., Lambrecht, J., Moeneclaey, B., Wang, C., and Yin, X. “Analog Demultiplexing 128-GBd Optical Receiver Integrating a Traveling-Wave TIA and a Current-Mode Sampling Front-End.”, *Journal of Lightwave Technology*, vol. 43, pp. 6250–6257, 2025.

Derevyanko, S., & Turitsyn, S., “Bit-Error Probability for Direct Detection of Optical RZ Signal Degraded by ASE Noise and Timing Jitter.” *Journal of Lightwave Technology*, 25, 638-643, 2007.

Dghais, W., Cunha, T., & Pedro, J. (2010). “Behavioral model for high-speed digital buffer/driver”. *Proc. INMMiC*, 110-113, 2010.

Dhruv, A., and Dubey, A., “Managing Software Provenance to Enhance Reproducibility in Computational Research,” *Computing in Science & Engineering*, vol. 25, no. 3, pp. 60-65, 2023.

Dionisio, R.P., Parca, G., Reis, C., & Teixeira, A.L. “Operational parameter optimisation of MZI-SOA using multi-objective genetic algorithms”. *Electronics Letters*, 47, 561-562, 2011.

Dixon-Declève, S., Balland, P. A., et al. “Industry 5.0: A Transformative Vision for Europe - Governing Systemic Transformations Towards a Sustainable Industry”. ESIR Policy Brief No. 3, European Commission, Directorate-General for Research and Innovation, Brussels, 2022.

Donati, A. “Roma Capitale: Smart Mobility per la mobilità sostenibile. Energia, Ambiente e Innovazione (ENEA),” Smart Future section, 2022.

D’Amico, A., Correia, B., London, E., Virgillito, E., Borraccini, G., Napoli, A., Curri, V., “Scalable and Disaggregated GGN Approximation Applied to a C+L+S Optical Network”, *Journal of Lightwave Technology*, vol. 40, no. 11, pp. 3499-3511, 2022.

D'Apuzzo, M., D'Arco, M., Liccardo, A., and Vadursi, M., "Analyzing DAC waveform distortion due to finite settling time", in Proc. IEEE Instrumentation & Measurement Technology Conference (I2MTC), pp. 1398-1403, 2010.

D'Apuzzo, M., D'Arco, M., Liccardo, A., and Vadursi, M., "Modeling DAC Output Waveforms", IEEE Transactions on Instrumentation and Measurement, vol. 59, no. 11, pp. 2854-2862, 2010.

El-Masri, D., Petrillo, F., Guéhéneuc, Y.-G., Hamou-Lhadj, A., and Bouziane, A., "A Systematic Literature Review on Automated Log Abstraction Techniques", Information and Software Technology, vol. 122, art. 106276, 2020.

Ellen MacArthur Foundation, "Towards the Circular Economy Vol. 1: An economic and business rationale for an accelerated transition," Ellen MacArthur Foundation, Cowes 2013.

Elwan, H. H., Poette, J., Cabon, B., "Simplified Chromatic Dispersion Model Applied to Ultrawide Optical Spectra for 60 GHz Radio-Over-Fiber Systems", Journal of Lightwave Technology, vol. 37, no. 19, pp. 5115-5121, 2019.

European Commission, "A new Circular Economy Action Plan - For a cleaner and more competitive Europe," European Commission, COM(2020)98 final, Brussels 2020.

European Commission, "Closing the loop - An EU action plan for the Circular Economy," COM(2015) 614 final, Brussels 2015.

European Commission, "Industry 5.0 - Research and Innovation for a Sustainable, Human-Centric and Resilient European Industry," Research & Innovation Policy Brief, Brussels 2021.

Fan, Q., Zhou, G., Gui, T., Lu, C., and Lau, A. P. T., "Advancing theoretical understanding and practical performance of signal processing for nonlinear optical communications through machine learning," Nature Communications, vol. 11, article 3694, 2020.

Faruk, M. S., Savory, S. J., "Measurement Informed Models and Digital Twins for Optical Fiber Communication Systems", Journal of Lightwave Technology, vol. 42, no. 3, pp. 1016-1030, 2024.

Fava, A., Centurelli, F., Monsurró, P., & Scotti, G., “A Compact Low-Power Chopper Low Noise Amplifier for High Density Neural Front-Ends,” *Sensors*, vol. 25, no. 4, art. 1157, 2025.

Feng, Z., Li, N., Li, W., He, P., Luo, M., Hu, Q., Huang, L., and Jiang, Y., “A Simplified Volterra Equalizer Based on System Characteristics for Direct Modulation Laser (DML)-Based Intensity Modulation and Direct Detection (IM/DD) Transmission Systems”, *Photonics*, 10(10), 1174, 2023.

Freire, P.J., Napoli, A., Spinnler, B., Costa, N., Turitsyn, S.K., Prilepsky, J.E., “Neural Networks-Based Equalizers for Coherent Optical Transmission: Caveats and Pitfalls”, *IEEE Journal of Selected Topics in Quantum Electronics*, 28(4), 7600223, 2022.

Freire, P., Srivallapanondh, S., Spinnler, B., Napoli, A., Costa, N., Prilepsky, J., & Turitsyn, S., “Computational Complexity Optimization of Neural Network-Based Equalizers in Digital Signal Processing: A Comprehensive Approach,” *Journal of Lightwave Technology*, vol. 42, pp. 4177–4201, 2024.

Fu, Y., Zhang, X., Hraimel, B., Liu, T., & Shen, D. “Mach-Zehnder: A Review of Bias Control Techniques for Mach-Zehnder Modulators in Photonic Analog Links”, *IEEE Microwave Magazine*, vol. 14, no. 7, pp. 102-107, 2013.

Fu, Y., Xin, H., Zhang, K., Bai, J., Bi, M., Miao, X., Li, L., Guo, Z., He, H., and Hu, W., “Impact of SOA-induced pattern effect on the filter requirements in vestigial sideband direct detected PAM4 transmission”, *Optics Express*, vol. 26, no. 23, pp. 30305–30312, 2018.

Ganjali, M., & Hosseini, S., “Effects of frequency chirping and finite extinction ratio of optical modulators in microwave photonic IFM receivers”, *Optics Communications*, 452, 380-386, 2019.

Gappmair, W., “Non-Data-Aided SNR Estimation for Bandlimited Optical Intensity Channels”, *Sensors*, vol. 23, no. 2, Art. 802, 2023.

Gappmair, W., “Novel Results on SNR Estimation for Bandlimited Optical Intensity Channels”, *Sensors*, vol. 24, no. 1, Art. 23, 2024.

Gaur, T., Mishra, P., Hegde, G., & Srinivas, T., “Modeling and Analysis of Device Orientation, Analog and Digital Performance of Electrode Design for High Speed Electro-Optic Modulator”, *Photonics*, 10(3), 301, 2023.

Ge, J., Zhang, J., Xu, S., Wang, Q., Lin, J., Hu, S., Lu, X., Ou, Z., Wang, S., Wang, T., Li, Y., Chen, J., Zhang, T., and Zhou, W., “82.5 GHz Photonic W-Band IM/DD PS-PAM4 Wireless Transmission over 300 m Based on Balanced and Lightweight DNN Equalizer Cascaded with Clustering Algorithm”, *Sensors*, vol. 25, 2025.

Geissdoerfer, M., Savaget, P., Bocken, N. M. P., Hultink, E. J., “The Circular Economy - A new sustainability paradigm?,” *Journal of Cleaner Production*, 143, 757-768, 2017.

Geng, Z., Kong, D., Rozental, V., Lowery, A., and Corcoran, B., “Optical sampling to enhance Nyquist-shaped signal detection under limited receiver bandwidth”, *Optics Express*, vol. 27, no. 17, pp. 24007–24017, 2019.

Ghazisaeidi, A., “A Theory of Nonlinear Interactions Between Signal and Amplified Spontaneous Emission Noise in Coherent Wavelength Division Multiplexed Systems”, *Journal of Lightwave Technology*, vol. 35(23), pp. 5150-5175, 2017.

Gheorghe, C., & Soica, A. “Revolutionizing urban mobility: A systematic review of AI, IoT, and predictive analytics in adaptive traffic control systems for road networks.” *Electronics*, 14(4), 719, 2025.

Ghosh, S., and Sahoo, B., “Closed-Form Expression for the Combined Effect of Offset, Gain, Timing, and Bandwidth Mismatch in Time-Interleaved ADCs Using Generalized Sampling”, *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 2007112, 2021.

Giffinger, R., Fertner, C., Kramar, H., Kalasek, R., Pichler-Milanović, N., & Meijers, E. J. “Smart Cities. Ranking of European Medium-Sized Cities”. Final Report. University of Copenhagen. 2007.

Ginovart, F., Simon, J.C., Valiente, I., “Gain recovery dynamics in semiconductor optical amplifier”. *Optics Communications*, 199, 111-115, 2001.

Giuliani, G., & D'Alessandro, D., "Noise analysis of conventional and gain-clamped semiconductor optical amplifiers." *Journal of Lightwave Technology*, 18, 1256-1263, 2000.

Glaessgen, E. H., and Stargel, D. S., "The Digital Twin Paradigm for Future NASA and U.S. Air Force Vehicles", *AIAA Paper 2012-1818*, 2012.

Grabowski, A., Gustavsson, J., and Larsson, A., "Large-Signal Equivalent Circuit for Datacom VCSELs - Including Intensity Noise," *Journal of Lightwave Technology*, vol. 40, pp. 7342-7350, 2022.

Gracias, J. S., Parnell, G. S., Specking, E., Pohl, E. A., & Buchanan, R. "Smart Cities - A Structured Literature Review." *Smart Cities*, 6(4),1719-1743, 2023.

Granot, E., Katz, G., "PAM-N—Fundamental Limits in Chromatic Dispersive-Uncompensated Channels", *Applied Sciences*, vol. 11, no. 6, Art. 2542, 2021.

Gu, S., Rong, G., Zhang, H., and Shen, H., "Logging Practices in Software Engineering: A Systematic Mapping Study", *IEEE Transactions on Software Engineering*, vol. 49, no. 2, pp. 902-923, 2023.

Gutiérrez, A., Brimont, A., Herrera, J., Aamer, M., Thomson, D., Gardes, F., Reed, G., Fédéli, J., & Sanchis, P., "Analytical Model for Calculating the Nonlinear Distortion in Silicon-Based Electro-Optic Mach-Zehnder Modulators", *Journal of Lightwave Technology*, 31(23), 3603-3613, 2013.

Gutiérrez-Castrejón, R., Saber, M., Alam, M., Xing, Z., El-Fiky, E., Ceballos-Herrera, D., Cavaliere, F., Vall-Llosera, G., Giorgi, L., Lessard, S., Brunner, R., & Plant, D., "Systematic Performance Comparison of (Duobinary)-PAM-2,4 Signaling under Light and Strong Opto-Electronic Bandwidth Conditions", *Photonics*,8(3),81, 2021.

Hayle, S., Hsu, H., Wang, C., Lu, H., Lu, J., Hsu, W., Chung, Y., Bai, Y., and Okram, K., "High-speed FSO-5G wireless communication system with enhanced loss compensation using high-power EDFA", *Scientific Reports*, vol. 15, 2025.

He, S., He, P., Chen, Z., Yang, T., Su, Y., and Lyu, M. R., "A Survey on Automated Log Analysis for Reliability Engineering", *ACM Computing Surveys*, vol. 54, no. 6, art. 130, pp. 1-37, 2021.

Heck, M., “Energy-efficient semiconductor optical amplifier with longitudinally varying confinement factor”, *Optics Letters*, 50 (2), 301–304, 2025.

Hossain, M., Böcherer, G., Rahman, T., Wettlin, T., Stojanović, N., Calabrò, S., and Pachnicke, S., “Probabilistic Shaping for High-Speed Unamplified IM/DD Systems With an O-Band EML”, *Journal of Lightwave Technology*, vol. 41, pp. 5373–5382, 2023.

Huang, X., Xie, F., Liu, S., Xu, H., Bai, C., & Qiao, Y. “Optimized equalization based on MAP detection for C-band bandwidth-limited IM/DD systems.” *Optics letters*, 49 13, 3644-3647, 2024.

Huang, H.-Y., Chen, X.-Y., and Kuo, T.-H., “A 10-GS/s NRZ/Mixing DAC With Switching-Glitch Compensation Achieving SFDR > 64/50 dBc Over the First/Second Nyquist Zone,” *IEEE Journal of Solid-State Circuits*, vol. 56, no. 10, pp. 3145-3156, 2021.

Hugar, N. R., Piruthiviraj, P., Maleeha, and Dhinesh, V., "Simulation and Analysis of Bit Error Rate in Optical Fiber Communication Using Optisystem", 2024 7th International Conference on Devices, Circuits, and Systems (ICDCS), pp. 67-71, 2024.

Hunter-Zinck, H., de Siqueira, A. F., Vásquez, V. N., Barnes, R., and Martinez, C. C., "Ten simple rules on writing clean and reliable open-source scientific software," *PLoS Computational Biology*, vol. 17, no. 11, e1009481, 2021.

Huseien, G., & Shah, K. “Potential Applications of 5G Network Technology for Climate Change Control: A Scoping Review of Singapore,” *Sustainability*, 13(17), 9720, 2021.

IEEE, “IEEE Standard for Ethernet—Amendment 10: Media Access Control Parameters, Physical Layers, and Management Parameters for 200 Gb/s and 400 Gb/s Operation,” *IEEE Std 802.3bs-2017*, 2017.

IEEE, “IEEE Standard for Ethernet,” *IEEE Std 802.3* (latest consolidated edition; RS-FEC/KP4 clauses for PAM4 Ethernet PHYs), IEEE, 2022.

Isaksson, M., Wisell, D., and Rönnow, D., “A comparative analysis of behavioral models for RF power amplifiers,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 54, no. 1, pp. 348-359, 2006.

Ismagilova, E., Hughes, L., Dwivedi, Y. K., & Raman, K. R. “Smart Cities: Advances in Research-An Information Systems Perspective.” *International Journal of Information Management*, 47, 201988-100, 2019.

ITU-T, “Digital test patterns for performance measurements on digital transmission equipment,” ITU-T Recommendation O.150, Oct. 1996.

Jang, S., Lee, J., Choi, Y., Kim, D., Kim, G., “Recent Advances in Ultrahigh-Speed Wireline Receivers With ADC-DSP-Based Equalizers”, *IEEE Open Journal of the Solid-State Circuits Society*, vol. 4, pp. 290–304, 2024.

Jauregui-Sánchez, Y., Clemente, P., Latorre-Carmona, P., Tajahuerce, E., and Lancis, J., “Signal-to-noise ratio of single-pixel cameras based on photodiodes”, *Applied Optics*, vol. 57, no. 7, pp. B67–B73, 2018.

Javaid, M., Haleem, A., Singh, R., Haq, M. I. U., “An integrated outlook of cyber-physical systems for Industry 4.0, Topical practices, architecture, and applications” *Green Technologies and Sustainability*, 1(1), 100001, 2023.

Ji, B., Zhang, X., Mumtaz, S., Han, C., Li, C., Wen, H., & Wang, D. “Survey on the Internet of Vehicles: Network Architectures and Applications,” *IEEE Communications Standards Magazine*, 4(1), 34-41, 2020.

Jiang, J., Yin, M., Yang, D., and Li, F., “Generation of high-speed PAM-4 signal with 3-bit DAC enabled by CRD-NS in optical interconnect,” *Optics Express*, vol. 31, no. 13, pp. 21153-21160, 2023.

Jones, D., Snider, C., Nassehi, A., Yon, J., and Hicks, B., “Characterising the Digital Twin: A systematic literature review”, *CIRP Journal of Manufacturing Science and Technology*, vol. 29, pp. 36-52, 2020.

Josten, A., Baeuerle, B., Bonjour, R., Heni, W., and Leuthold, J., “Optical Transmitters without Driver Amplifiers—Optimal Operation Conditions,” *Applied Sciences*, vol. 8, no. 9, art. 1686, 2018.

Kaje, K., Al-Qadi, M., and Hui, R., “Digital Compensation of SOA-Induced Nonlinearities in Field-Modulated Direct-Detection Systems”, *IEEE Photonics Journal*, 13(6), 7200405, 2021.

Kaluarachchi, Y. “Implementing data-driven smart city applications for future cities,” *Smart Cities*, 5(2), 350-367, 2022.

Karanov, B., Chagnon, M., Thouin, F., et al., “End-to-End Deep Learning of Optical Communications,” *Journal of Lightwave Technology*, vol. 36, pp. 4843-4851, 2018.

Karar, A., Falou, A., Barakat, J., Gürkan, Z., & Zhong, K. “Recent Advances in Coherent Optical Communications for Short-Reach: Phase Retrieval Methods,” *Photonics*, 10(3), 308, 2023.

Keiser, G., “Optical Fiber Communications”, McGraw-Hill Education, 5th ed., 2013.

Kennedy, M. C., and O’Hagan, A., “Bayesian calibration of computer models”, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 3, pp. 425-464, 2001.

Kester, W. (Ed.), "The Data Conversion Handbook", Newnes/Elsevier, ISBN: 978-0-7506-7841-4, 2005. also published by Analog Devices, Inc. under the title “Analog-Digital Conversion,” ISBN 0-916550-27-3. 2004

Khan, A., Turowski, K., “A Preliminary Study on Industry 4.0,” *Journal of Industrial and Intelligent Information*, 4(3), 2016.

Khan, L.U., Yaqoob, I., Tran, N.H., Kazmi, S., Dang, T.N., & Hong, C. “Edge-Computing-Enabled Smart Cities: A Comprehensive Survey,” *IEEE Internet of Things Journal*, 7, 10200-10232, 2020.

Khan, S., et al. “The Role of Industry 5.0 in Achieving Circular Economy Goals: Integrating Smart Technologies for Sustainable Supply Chains.” *TEM Journal*, 14(4), 3144-3154, 2025.

Khan, F. N., Fan, Q., Lu, C., & Lau, A. P. T., “Machine Learning Methods for Optical Communication Systems and Networks”, Optical Fiber Telecommunications VII, pp. 921-978, 2020.

Kharraz, O. M., Supa’at, A. S. M., Atieh, A., Zamzuri, A.K., and Mahdi, M. A., “Enhanced gain saturation model of non-linear semiconductor optical amplifiers”, IET Optoelectronics, vol. 12, no. 6, pp. 263–268, 2018.

Kim, M. H., Yu, B. M., and Choi, W. Y., “A Mach-Zehnder Modulator Bias Controller Based on OMA and Average Power Monitoring,” IEEE Photonics Technology Letters, vol. 29, no. 23, pp. 2043-2046, 2017.

Kincaid, P., Andriolli, N., Contestabile, G., & De Marinis, L., “Addressing optical modulator non-linearities for photonic neural networks”, Communications Engineering, 4,58, 2025.

Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., Inman, D. J., “1D Convolutional Neural Networks and Applications: A Survey”, Mechanical Systems and Signal Processing, vol. 151, Art. no. 107398, 2021.

Kirchherr, J., Reike, D., Hekkert, M., “Conceptualizing the circular economy: An analysis of 114 definitions,” Resources, Conservation and Recycling, 127, 221-232, 2017.

Klopries, H., Torres, D., & Schwung, A. “Synthetic time series dataset generation for unsupervised autoencoders.” 2022 IEEE 27th International Conference on Emerging Technologies and Factory Automation (ETFA), 1-8. 2022.

Kottke, C., Schmidt, C., Jungnickel, V., & Freund, R. “Performance of Bandwidth Extension Techniques for High-Speed Short-Range IM/DD Links”, Journal of Lightwave Technology, vol. 37, pp. 665-672, 2019.

Kraemer, R., Düzgöl, O., Li, S., and Calabretta, N., “Data-Driven SOA Parameter Discovery and Optimization Using Bayesian Machine Learning With a Parzen Estimator Surrogate”, Journal of Lightwave Technology, vol. 42, pp. 721–731, 2024.

Kreuzer, T., Papapetrou, P., Zdravkovic, J., “Artificial intelligence in digital twins - A systematic literature review”, Data & Knowledge Engineering, vol. 151, art. 102304, 2024.

Kritzing, W., Karner, M., Traar, G., Henjes, J., and Sih, W., “Digital Twin in manufacturing: A categorical literature review and classification”, IFAC-PapersOnLine, vol. 51, no. 11, pp. 1016-1022, 2018.

Kumar, A., Batra, N., Mudgal, A., & Yadav, A. (2024). “Navigating urban mobility: A review of AI-driven traffic flow management in smart cities,” In Proc. 11th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO), 2024.

Kumari, A., Bhatt, D., “Advanced System Analysis and Survey on the GPS Receiver Front End”, IEEE Access, vol. 10, pp. 24611–24626, 2022.

Kuntze, S. B., Zilkie, A. J., Pavel, L., and Aitchison, J. S., “Nonlinear State-Space Model of Semiconductor Optical Amplifiers With Gain Compression for System Design and Analysis”, Journal of Lightwave Technology, vol. 26, no. 14, pp. 2274–2281, 2008.

Kurosawa, N., Kobayashi, H., Maruyama, K., Sugawara, H., & Kobayashi, K. “Sampling clock jitter effects in digital-to-analog converters”.Measurement, 31(3), 187-199, 2002.

Köhler, C.A., Ulianych, D., Grün, S., Decker, S. and Denker, M., “Facilitating the Sharing of Electrophysiology Data Analysis Results Through In-Depth Provenance Capture”, eNeuro, 2024.

Le, Q., Feng, K., and Kämpfe, T., “Analysis of Calibration and De-embedding for On-Wafer Characterization up to 170 GHz in 22-nm FDSOI CMOS Technology,” 2025 IEEE 24th Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems (SiRF), 83-85, 2025.

Leipzig, J., Nüst, D., Hoyt, C. Tapley, Ram, K., and Greenberg, J., “The role of metadata in reproducible computational research”, Patterns, vol. 2, no. 9, 100322, 2021.

Lennard, S., Barbosa, F.A., & Ferreira, F.M., “Zero-Shot ML Equalization in Coherent Optical Transmission: A First Experimental Demonstration”. Journal of Lightwave Technology, 43, 6194-6202, 2025.

Li X. & Zhou L., “A survey of high-speed and high-resolution current-steering DACs,” Journal of Semiconductors, vol. 41, no. 11, art. 111404, 2020.

Li, J., An, S. H., Zhu, Q., Li, X., and Su, Y., “VSB Modified Duobinary PAM4 Signal Transmission in an IM/DD System With Mitigated Image Interference”, *IEEE Photonics Technology Letters*, vol. 32, pp. 363-366, 2020.

Li, Z., Yi, L., and Hu, W., “Symmetric 40-Gb/s TWDM-PON with 51-dB loss budget by using a single SOA as preamplifier, booster and format converter in ONU”, *Optics Express*, vol. 22, no. 20, pp. 24398-24404, 2014.

Li, C., and Mahadevan, S., “Role of calibration, validation, and relevance in multi-level uncertainty integration”, *Reliability Engineering & System Safety*, vol. 148, pp. 32-43, 2016.

Li, J., Su, X., Ye, T., Nakashima, H., Hoshida, T., and Tao, Z., “Characterization of Nonlinear Distortion in Intensity Modulation and Direct Detection Systems”, *Journal of Lightwave Technology*, vol. 41(11), pp. 3513-3521, 2023.

Li, Z., Li, Y., Luo, S., Yin, F., Wang, Y., and Song, Y., “SOA Amplified 100 Gb/s/ λ PAM-4 TDM-PON Supporting PR-30 Power Budget with >18 dB Dynamic Range”, *Micromachines*, vol. 13, no. 3, art. 342, 2022.

Li, D., Geng, L., Maloberti, F., and Svelto, F., “Overcoming the Transimpedance Limit: A Tutorial on Design of Low-Noise TIA”, *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 69, pp. 2648–2653, 2022.

Li, H., Hsu, C., Sharma, J., Jaussi, J., and Balamurugan, G. “A 100-Gb/s PAM-4 Optical Receiver With 2-Tap FFE and 2-Tap Direct-Feedback DFE in 28-nm CMOS.”, *IEEE Journal of Solid-State Circuits*, vol. 57, pp. 44–53, 2022.

Li, K., Liu, S., Ruan, X., Thomson, D., Hong, Y., Yang, F., Zhang, L., Lacava, C., Meng, F., Zhang, W., Petropoulos, P., Zhang, F., and Reed, G., “Co-design of a differential transimpedance amplifier and balanced photodetector for a sub-pJ/bit silicon photonics receiver”, *Optics Express*, vol. 28, no. 9, pp. 14038–14054, 2020.

Li, Q., Chen, B., Lei, Y., and Liang, Z., “On the performance and fiber nonlinearity modeling of learned digital back-propagation in the presence of DAC and ADC quantization noise,” *Optics Express*, vol. 33, no. 4, pp. 7962–7977, 2025.

Li, S., Li, H., Dai, W., Li, C., Zou, J., & Xiong, H., “Learned Progressive Image Compression With Dead-Zone Quantizers”, *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, pp. 2962–2978, 2023.

Li, G., Li, Z., Ha, Y., Hu, F., Zhang, J., Chi, N., “Performance Assessments of Joint Linear and Nonlinear Pre-Equalization Schemes in Next Generation IM/DD PON”, *Journal of Lightwave Technology*, vol. 40, pp. 5478–5489, 2022.

Li, Y., Wei, Z., Liu, L., Fu, H., Yu, C., “Mitigated multipath interference in PAM4 IM/DD optical links using instantaneous mean intensity addressing,” *Optics Express* vol. Vol. 33, Issue 9, pp. 18855-18865, 2025.

Li, Z., Liu, F., Yang, W., Peng, S., Zhou, J., “A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects”, *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 12, pp. 6999–7019, 2022.

Liang, E., and Kahn, J., “Geometric Shaping for Distortion-Limited Intensity Modulation/Direct Detection Data Center Links”, *IEEE Photonics Journal*, vol. 15, pp. 1–17, 2023.

Liao, T., Xue, L., Huang, L., Hu, W., & Yi, L., “Training data generation and validation for a neural network-based equalizer,” *Optics Letters*, vol. 45, no. 18, pp. 5113–5116, 2020.

Liu, L., Chen, C., Pei, Q., Maharjan, S., & Zhang, Y. “Vehicular Edge Computing and Networking: A Survey,” *Mobile Networks and Applications*, 26, 1145-1168, 2021.

Liu, X., Zhang, Y., Chen, Y., Liu, Y., Cai, M., Qiu, Q., Fu, M., Yi, L., Hu, W., Zhuge, Q., “Digital twin modeling and controlling of optical power evolution enabling autonomous-driving optical networks: a Bayesian approach”, *Advanced Photonics*, vol. 6, no. 2, 026006, 2024.

Liu, S., “Research on high-speed digital optical signal jitter measurement technology based on clock recovery algorithm using eye diagram opening area”, *Heliyon*, vol. 10, Art. e35756, 2024.

Liu, W., Zhang, T., Zhang, H., Tang, H., and Zhang, L., “A glitch energy reduction method with low complexity dynamic element matching for data converters”, *IEICE Electronics Express*, vol. 20, 20220529, 2023.

Liu, Y., Yong, S., Gao, H., Hinaga, S., Padilla, D., Yanagawa, D., Drewniak, J., and Khilkevich, V., “S-Parameter De-Embedding Error Estimation Based on the Statistical Circuit Models of Fixtures,” *IEEE Transactions on Electromagnetic Compatibility*, 62, 1459-1467, 2020.

Liu, S., Ming, J., Song, J., Mu, Y., He, H., Wang, R., Gao, R., Li, Z., Xin, X., Dong, Z., “Least-squares-assisted BCJR algorithm for severe ISI mitigation in bandwidth-limited IM/DD optical interconnection”, *Optics Express*, vol. 33, no. 12, pp. 24545–24552, 2025.

Liu, X., Li, Z., Wen, H., Miao, M., Wang, Y., and Wang, Z., “A PAM4 transceiver design scheme with threshold adaptive and tap adaptive,” *EURASIP Journal on Advances in Signal Processing*, 2023, 70, 2023.

Liu, X., Hu, C., Cai, Y., and Zhao, J., “Exponential-Memory-Polynomial-Based Nonlinear Equalizer for C-Band IM/DD Systems,” *IEEE Photonics Technology Letters*, 36, 1261–1264, 2024.

Liu, X., Liu, Y., Hu, C., and Zhao, J., “Experimental demonstration of low-complexity trigonometric-memory-polynomial decision-feedback equalizer for nonlinear compensation in IM/DD optical systems”, *Optics Express*, vol. 32, no. 3, pp. 3329–3341, 2024.

Locatelli, F., Christodouloupoulos, K., Svaluto Moreolo, M., Fàbrega, J. M., and Spadaro, S., “Machine Learning-Based In-Band OSNR Estimation From Optical Spectra”, *IEEE Photonics Technology Letters*, vol. 31, no. 24, pp. 1929-1932, 2019.

Loia, F., Perillo, C., & Gravili, G. “Unleashing the power of digital twin and big data as a new frontier for smart mobility: An ecosystem perspective,” *Big Data Research*, vol 43, 2025.

Long, J., Wang, C., Wu, Y., Sang, B., Bian, C., Yang, X., Zhang, L., Chen, Y., Zhang, Q., Wang, Y., Li, Y., Zhou, W., Wang, K., Liu, B., Shen, L., and Yu, J., “Digital Frequency-

Domain MIMO Equalizer Enabling Six-LP-Mode Strong-Coupling IM/DD MDM Optical Transmission System”, *Sensors*, vol. 25, 2025.

Lorenz, A., Matthus, C., & Ellinger, F. “Microstrip-Based Anti-Aliasing Filters for Beyond-50 GHz-Bandwidth Mix-Filter-Sample Analog-to-Digital-Converters,” *IEEE Access*, 13, 2025.

Lu, D., Zhou, X., Yang, Y., Huo, J., Yuan, J., Long, K., Yu, C., Lau, A., and Lu, C., “Theoretical analysis of PAM-N and M-QAM BER computation with single-sideband signal”, *Science China Information Sciences*, vol. 64, 182312, 2021.

Mahmoud, H.A., & Arslan, H. “Error Vector Magnitude to SNR Conversion for Nondata-Aided Receivers”. *IEEE Transactions on Wireless Communications*. 8(5), 2009.

Mangiaracina, R., Perego, A., Salvadori, G., & Tumino, A. “A comprehensive view of intelligent transport systems for urban smart mobility.” *International Journal of Logistics Research and Applications*, 20(1), 39-52, 2017.

Marco, D., and Neuhoff, D. L., “The validity of the additive noise model for uniform scalar quantizers,” *IEEE Transactions on Information Theory*, vol. 51, no. 5, pp. 1739-1755, 2005.

Merz, P. T., and Shirts, M. R., "Testing for physical validity in molecular simulations," *PLoS ONE*, vol. 13, no. 7, e0202764, 2018.

Meslener, G. J., “Chromatic dispersion induced distortion of modulated monochromatic light employing direct detection”, *IEEE Journal of Quantum Electronics*, vol. 20, no. 10, pp. 1208-1216, 1984.

Meuer, C., Kim, J., Laemmlin, M., Liebich, S., Capua, A., Eisenstein, G., Kovsh, A., Mikhrin, S., Krestnikov, I., & Bimberg, D., “Static gain saturation in quantum-dot semiconductor optical amplifiers”, *Optics Express*, vol. 16, no. 11, pp. 8269–8279, 2008.

Michaeli, L., and Šaliga, J., “Error models of the analog to digital converters,” *Measurement Science Review*, vol. 14, no. 2, pp. 62-77, 2014.

Middleton, D., "Non-Gaussian Noise Models in Signal Processing for Telecommunications: New Methods and Results for Class A and Class B Noise Models," IEEE Transactions on Information Theory, vol. 45, no. 4, pp. 1129-1149, 1999.

Mienye, I. D., Swart, T. G., Obaido, G., Jordan, M., Ilono, P., "Deep Convolutional Neural Networks in Medical Image Analysis: A Review", Information, vol. 16, no. 3, Art. no. 195, 2025.

Miqueu, P., Belvèze, F., Brossier, J., and Ros, L., "Parameters setting based on bit error probability formula for an amplitude modulated, baseband, communication system affected by timing jitter and frequency selective channel", AEU - International Journal of Electronics and Communications, 178, art. 155288, 2024.

Mishra, P., & Singh, G. "Energy Management Systems in Sustainable Smart Cities Based on the Internet of Energy: A Technical Review," Energies, 16(19), 6903, 2023.

Mitieka, D., Luke, R., Twinomurinzi, H., & Mageto, J. "Smart mobility in urban areas: A bibliometric review and research agenda." Sustainability, 15(8), 6754, 2023.

Moon, J., and Kim, B., "Enhanced Hammerstein Behavioral Model for Broadband Wireless Transmitters," IEEE Transactions on Microwave Theory and Techniques, vol. 59, no. 4, pp. 924-933, 2011.

Morris, T. P., White, I. R., and Crowther, M. J., "Using simulation studies to evaluate statistical methods," Statistics in Medicine, vol. 38, no. 11, pp. 2074-2102, 2019.

Mu, Y., He, H., Ming, J., Song, J., Xu, Q., Gao, R., Li, Z., et al., "End-to-End Learning Strategy with Channel-Aided Polar Autoencoder in IM/DD Optical Interconnection," Optics Express, vol. 32, no. 22, pp. 39727-39733, 2024.

Munhoz, P., da Costa Dias, F., Chinelli, C., Guedes, A., Santos, J., da Silveira e Silva, W., & Soares, C. "Smart mobility: The main drivers for increasing the intelligence of urban mobility." Sustainability, 12(24), 10675, 2020.

Murao, T., Ushida, J., Takahashi, H., Tokushima, M., Shiina, A., & Horikawa, T., "Understanding Inter-Arm Imbalance in Mach-Zehnder Silicon Optical Modulators: Impact on OMA and Frequency Chirp", Journal of Lightwave Technology, 40(15), 5171-5189, 2022.

Murphy, S., Jamali, F., Townsend, P., and Antony, C., “High Dynamic Range 100G PON Enabled by SOA Preamplifier and Recurrent Neural Networks”, *Journal of Lightwave Technology*, vol. 41(11), pp. 3522–3532, 2023.

Musumeci, F., Rottondi, C., Nag, A., Macaluso, I., Zibar, D., Ruffini, M., “An Overview on Application of Machine Learning Techniques in Optical Networks”, *IEEE Communications Surveys & Tutorials*, vol. 21, pp. 1383-1408, 2019.

Muñiz-Cánovas, L. M., Barmenkov, Y. O., Kiryanov, A. V., Cruz, J. L., Andrés, M.V., “ASE narrow-band noise pulsing in erbium-doped fiber amplifier and its effect on self-phase modulation”, *Optics Express*, vol. 27, no. 6, pp. 8520-8528, 2019.

Mölder, F., et al., “Sustainable data analysis with Snakemake”, *F1000Research*, vol. 10, Art. 33, 2021.

Nadimi Goki, P., Imran, M., Fresi, F., Cavaliere, F., and Potì, L., “Lossless ROADM by Exploiting low gain SOAs in fronthaul network”, in *Proc. 2019 OptoElectronics and Communications Conference (OECC) / 2019 International Conference on Photonics in Switching and Computing (PSC)*, 2019.

Nadimi Goki, P., Imran, M., Porzi, C., Toccafondo, V., Fresi, F., Cavaliere, F., and Potì, L., “Lossless WDM PON Photonic Integrated Receivers Including SOAs”, *Applied Sciences*, vol. 9, no. 12, Art. 2457, 2019.

Napoli, A., Mezghanni, M. M., Calabrò, S., Palmer, R., Saathoff, G., and Spinnler, B., “Digital Predistortion Techniques for Finite Extinction Ratio IQ Mach-Zehnder Modulators,” *Journal of Lightwave Technology*, vol. 35, no. 19, pp. 4289-4296, 2017.

Nasir, A., Durrani, S., Mehrpouyan, H., Blostein, S., Kennedy, R., “Timing and carrier synchronization in wireless communication systems: a survey and classification of research in the last 5 years”, *EURASIP Journal on Wireless Communications and Networking*, 180, 2016

Nayeri, S., Sazvar, Z., & Babae Tirkolaee, E. B. “Towards Waste Management 5.0: A Nexus Between Circular Economy and Industry 5.0 Dimensions.” *Waste Management*, 203, 114830, 2025.

Nazarathy, M., & Tomkos, I., “‘Perfect’ PAM4 serial digital-optical conversion,” *IEEE Photonics Technology Letters*, vol. 33, no. 9, pp. 475-478, 2021.

Nazarathy, M., & Tomkos, I., “All-Analytic Statistical Modeling of Constellations in (Optical) Transmission Systems Driven by High-Speed Electronic Digital to Analog Converters Part I: DAC Mismatch Statistics, Metrics, Symmetries, Error Vector Magnitude”, *Photonics*, vol. 11, no. 8, Art. 747, 2024.

Niu, Z., Yang, H., Li, L., Shi, M., Xu, G., Hu, W., & Yi, L., "Learnable digital signal processing benchmark for linearity compensation in optical fiber communications," *Light: Science & Applications*, 2024.

Noe, R., “Extension for ‘Consistent Optical and Electrical Noise Figure’,” *Journal of Lightwave Technology*, vol. 42, no. 10, pp. 3540-3545, 2024.

Nussio, F. “Development of mobility in Rome with REACT-EU structural funds: from new LEZ and traffic monitoring to Rome MaaS”. Presentation at Barcelona Smart City Expo 2023 (PON Metro / REACT-EU), 2023.

Oladimeji, D., Gupta, K., Kose, N. A., Gundogan, K., Ge, L., & Liang, F. “Smart transportation: An overview of technologies and applications,” *Sensors*, 23(8), 3880, 2023.

Omidi, A., Banawan, M., Weckenmann, E., Paquin, B., Geravand, A., Zheng, Z., Shi, W., Zeng, M., and Rusch, L., “Auto-Encoder Optimized PAM IM/DD Transceivers for Amplified Fiber Links”, *Journal of Lightwave Technology*, vol. 43, pp. 547–556, 2024.

Oppenheim, A. V., and Schafer, R. W., “Discrete-Time Signal Processing,” 3rd ed., Pearson, 2010.

Oveisi, M., & Heydari, P. “A Study of BER and EVM Degradation in Digital Modulation Schemes Due to PLL Jitter and Communication-Link Noise”. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 69(8), 3402 - 3415, 2022.

Pang, X., Ozolins, O., Lin, R., Zhang, L., Udalcovs, A., Xue, L., Schatz, R., Westergren, U., Xiao, S., Hu, W., Jacobsen, G., Popov, S., and Chen, J., “200-Gb/s-Lane IM-DD Technologies for Short-Reach Optical Interconnects,” *Journal of Lightwave Technology*, vol. 38, no. 2, pp. 492-503, 2020.

Patel, D., Sharif-Bakhtiar, A., and Carusone, T., “A 112-Gb/s —8.2-dBm Sensitivity 4-PAM Linear TIA in 16-nm CMOS With Co-Packaged Photodiodes”, *IEEE Journal of Solid-State Circuits*, vol. 58, pp. 771–784, 2023.

Pavan, S., Manivannan, S., & Basavaraj, N. “Analysis and Design of Wideband Filtering ADCs Using Continuous-Time Pipelining,” *IEEE Journal of Solid-State Circuits*, 59, 2024.

Petit-Boix, A., Leipold, S., “Circular economy in cities: Reviewing how environmental research aligns with local practices,” *Journal of Cleaner Production*, 195 (2018), 1270-1281, 2018.

Plabst, D., Prinz, T., Wiegart, T., Calabrò, S., Rahman, T., Stojanović, N., Hanik, N., and Kramer, G., “Achievable Rates for Short-Reach Fiber-Optic Channels With Direct Detection”, *Journal of Lightwave Technology*, vol. 40, pp. 3602–3613, 2021.

Proakis, J. G., and Salehi, M., “Digital Communications,” 5th ed., McGraw-Hill, 2007.

Qin, H., & Xiao, X., “Effects of fiber nonlinearity on error vector magnitude and bit error ratio for advanced modulation formats”, *Optical Engineering*, vol. 57, 2018.

Rasch, M. et al., “Hardware-aware training for large-scale and diverse deep learning inference workloads using in-memory computing-based accelerators,” *Nature Communications*, vol. 14, Art. 967, 2023.

Rathore, M. M., Shah, S. A., Shukla, Dh., Bentafat, E., BAKIRAS, S., “The Role of AI, Machine Learning, and Big Data in Digital Twinning: A Systematic Literature Review, Challenges, and Opportunities”, *IEEE Access*, vol. 9, pp. 32030-32052, 2021.

Rebola, J.L., and Cartaxo, A.V.T., “Q-factor estimation and impact of spontaneous-spontaneous beat noise on the performance of optically preamplified systems with arbitrary optical filtering”, *Journal of Lightwave Technology*, vol. 21(1), pp. 87–95, 2003.

Rejeb, A., Rejeb, K., Keogh, J. G., & Süle, E. “When Industry 5.0 Meets the Circular Economy: A Systematic Literature Review.” *Circular Economy and Sustainability*, 5(4), 2621-2652, 2025.

Reza, A., Troncoso-Costas, M., Browning, C., O’duill, S., and Barry, L., “Mitigation of SOA-Induced Nonlinearities With Recurrent Neural Networks in 75 Gbit/s/ λ PAM-4 IM/DD WDM-PON Transmission Systems”, *Journal of Lightwave Technology*, vol. 41, pp. 3967–3975, 2023.

Reza, A. G., Venkatasubramani, L. N., and Barry, L. P., “112.5 Gbit/s long reach passive optical network with over 31 dB power budget enabled by semiconductor optical amplifiers”, *Scientific Reports*, vol. 15, art. 27731, 2025.

Ribeiro, P., Dias, G., & Pereira, P. “Transport systems and mobility for smart cities.” *Applied System Innovation*, 4(3), 61, 2021.

Rizzelli, G., Torres-Ferrera, P., Forghieri, F., & Gaudino, R., “An Analytical Model for Performance Estimation in Modern High-Capacity IMDD Systems”, *Journal of Lightwave Technology*, 42(5), 1443-1452, 2024.

Robitaille, T. P., et al. (Astropy Collaboration), “Astropy: A community Python package for astronomy”, *Astronomy & Astrophysics*, vol. 558, A33, 2013.

Rodenbeck, C., Martinez, M., Beun, J., Silva-Martínez, J., Karsilayan, A., & Liechty, R., “When Less Is More ... Few Bit ADCs in RF Systems”, *IEEE Access*, vol. 7, pp. 12035–12046, 2019.

Roma Capitale. “Roma Smart City - Il Piano,” Programmatic document of Roma Capitale, 2021.

Roma Capitale. “Verso un futuro a zero emissioni - Il Progetto UPPER e le sue misure,” Institutional magazine, issue on the UPPER project and climate-neutrality commitments, 2025.

Roma Servizi per la Mobilità & Almaviva. “Roma Data Driven” (workshop/video session on data-driven mobility and smart-city services in Rome), 2024-2025.

Romanova, A., & Barzdėnas, “On Noise Modeling of Capacitive Feedback Transimpedance Amplifiers in CMOS,” *Applied Sciences*, vol. 12, no. 19, art. 10186, 2022.

Roumpos, I., De Marinis, L., Kirtas, M., Passalis, N., Tefas, A., Contestabile, G., Pleros, N., Moralis-Pegios, M., & Vysokinos, K., “High-Performance End-to-End Deep Learning IM/DD Link Using Optics-Informed Neural Networks”, *Optics Express*, vol. 31, no. 12, pp. 20068-20079, 2023.

Safarnejadian, A., Shi, W., Rusch, L. A., and Zeng, M., “Frequency-Dependent Variations in DAC ENOB: Predicting Accurate High-Speed Eye-Diagrams,” *IEEE Photonics Technology Letters*, vol. 37, no. 22, pp. 1297-1300, 2025.

Saif, W. S., Esmail, M. A., Ragheb, A. M., Alshaw, T. A., Alshebeili, S. A., “Machine Learning Techniques for Optical Performance Monitoring and Modulation Format Identification: A Survey”, *IEEE Communications Surveys & Tutorials*, Vol. 22, Issue: 4, pp. 2839 - 2882, 2020)

Saleh, B. E. A., and Teich, M. C., "Fundamentals of Photonics", Wiley, 3rd ed, 2019.

Sandve, G. K., Nekrutenko, A., Taylor, J., and Hovig, E., “Ten Simple Rules for Reproducible Computational Research”, *PLoS Computational Biology*, vol. 9, no. 10, e1003285, 2013.

Sang, B., Zhou, W., Tan, Y., Kong, M., Wang, C., Wang, M., Zhao, L., Zhang, J., & Yu, J., “Low Complexity Neural Network Equalization Based on Multi-Symbol Output Technique for 200+ Gbps IM/DD Short Reach Optical System,” *Journal of Lightwave Technology*, vol. 40, pp. 2890–2900, 2022.

Sarmiento, S., Mendiñeta, J., Altabás, J., Spadaro, S., Shinada, S., Furukawa, H., Olmos, J., Lázaro, J., and Wada, N., “Optical Power Budget Enhancement in 50–90 Gb/s IM-DD PONs With NOMA-CAP and SOA-Based Amplification”, *IEEE Photonics Technology Letters*, vol. 32, pp. 608–611, 2020.

Sasai, T., Nakamura, M., Yamazaki, E., Matsushita, A., Okamoto, S., Horikoshi, K., and Kisaka, Y., “Wiener-Hammerstein model and its learning for nonlinear digital predistortion of optical transmitters,” *Optics Express*, vol. 28, no. 21, pp. 30952-30963, 2020.

Satka, Z., Ashjaei, M., Fotouhi, H., Daneshtalab, M., Sjödin, M., & Mubeen, S. “A comprehensive systematic review of integration of time sensitive networking and 5G communication,” *Journal of Systems Architecture*, 138, 102852, 2023.

Saux, B., Borgmans, J., Raman, J., & Rombouts, P. “A 3.5 GS/s 1-1 MASH VCO ADC With Second-Order Noise Shaping,” *IEEE Transactions on Circuits and Systems I: Regular Papers*, 72, 2024.

Schmidt, C., Kottke, C., Freund, R., Gerfers, F., and Jungnickel, V. “Digital-to-analog converters for high-speed optical communications using frequency interleaving: impairments and characteristics”, *Optics Express*, vol. 26, no. 6, pp. 6758-6770, 2018.

Schmidt, C., Figueroa, J., Cousseau, J., & Tonello, A., “Pilot-Based TI-ADC Mismatch Error Calibration for IR-UWB Receivers”, *IEEE Access*, vol. 7, pp. 74340–74350, 2019.

Schoukens, J., Godfrey, K., & Schoukens, M., “Nonparametric Data-Driven Modeling of Linear Systems: Estimating the Frequency Response and Impulse Response Function.” *IEEE Control Systems*, 38, 49-88, 2018.

Schwamb, M. E., Kubica, J., Jurić, M., Oldag, D., West, M., DeLucchi, M., and Holman, M. J., “Controlling Randomization in Astronomy Simulations”, *Research Notes of the AAS*, vol. 8, no. 1, Article 25, 2024.

Schädler, M., Böcherer, G., & Pachnicke, S., “Soft-Demapping for Short Reach Optical Communication: A Comparison of Deep Neural Networks and Volterra Series,” *Journal of Lightwave Technology*, vol. 39, pp. 3095–3105, 2021.

Secondini, M., and Forestieri, E., “Direct Detection of Bipolar Pulse Amplitude Modulation”, *Journal of Lightwave Technology*, vol. 38, no. 21, pp. 5981-5990, 2020.

Secondini, M., Forestieri, E., Fresi, F., Poti, L., Catuogno, T., and Cavaliere, F., “Bipolar Pulse Amplitude Modulation with Direct Detection”, in *45th European Conference on Optical Communication (ECOC 2019)*, 2019.

Semeraro, C., Lezoche, M., Panetto, H., and Dassisti, M., “Digital twin paradigm: A systematic literature review”, *Computers in Industry*, vol. 130, 103469, 2021. DOI: 10.1016/j.compind.2021.103469.

Sena, M., Erkilinc, M.S., Dippon, T., Shariati, B., Emmerich, R., Fischer, J.K., “Bayesian Optimization for Nonlinear System Identification and Pre-Distortion in Cognitive Transmitters”. *Journal of Lightwave Technology*, 39(15), 5008-5020, 2021.

Sequeira, D., Ruiz, M., Costa, N., Napoli, A., Pedro, J., and Velasco, L., “OCATA: a deep-learning-based digital twin for the optical time domain analyzer”, *Journal of Optical Communications and Networking*, vol. 15, no. 2, pp. 87-97, 2023.

Shafique, M. N., Adeel, U., & Rashid, A. “The Synergy Between Industry 5.0 and Circular Economy for Sustainable Performance in the Chinese Manufacturing Industry.” *Sustainability*, 16(22), 9952, 2024.

Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., and de Freitas, N., “Taking the human out of the loop: A review of Bayesian optimization”, *Proceedings of the IEEE*, vol. 104, no. 1, pp. 148-175, 2016.

Shehab, M., Kassem, I., Kutty, A., Kucukvar, M., Onat, N., & Khattab, T. “5G Networks Towards Smart and Sustainable Cities: A Review of Recent Developments, Applications and Future Perspectives.” *IEEE Access*, 10, 2987-3006, 2022.

Sheng, W., Liu, C., Xiao, J., Sun, L., Cai, Y., Fu, H., Li, Q., & Liu, G., “Complex-valued recurrent neural network equalizer with low complexity for a 120-Gbps 50-km optical PAM-4 IM/DD system,” *Optics Express*, vol. 32, no. 16, pp. 27624–27634, 2024.

Shi, B., Zhou, Y., Sun, H., Nguyen, T., & Schutt-Ainé, J., “Statistical Method for Eye Diagram Simulation in High-Speed Link Nonlinear System Applications,” *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 14, no. 11, pp. 2050-2061, 2024.

Singh, J., Sandmann, A., Ahrens, A., Lochmann, S., “Joint pre- and post-equalization using optical multi-level signaling”, *The Journal of Engineering*, Vol. 2022, Issue4, pp. 442-446, 2022.

Smith, B. P., Farhood, A., Hunt, A., Kschischang, F. R., and Lodge, J., “Staircase Codes: FEC for 100 Gb/s OTN,” *Journal of Lightwave Technology*, vol. 30, no. 1, pp. 110-117, 2012.

Sobhanan, A., Anthur, A., O'Duill, S., Pelusi, M., Namiki, S., Barry, L., Venkitesh, D., & Agrawal, G. P. "Semiconductor optical amplifiers: recent advances and applications", *Advances in Optics and Photonics*, vol. 14, no. 3, pp. 571-651, 2022.

Solaimani, R., Nizamani, A., Civelli, S., Nadimi Goki, P., Cavaliere, F., and Potì, L., "Experimental Demonstration of Convolutional Neural Network Equalization for BPAM in IM-DD Systems", *Advanced Photonics Congress, SPPCom 2025*, paper SpTh1F.2, 2025.

Sozański, K., "Overview of Signal Processing Problems in Power Electronic Control Circuits", *Energies*, 16(12), 4774, 2023.

Stojanović, N., Karinou, F., Qiang, Z., and Prodaniuc, C., "Volterra and Wiener Equalizers for Short-Reach 100G PAM-4 Applications," *Journal of Lightwave Technology*, 35, 4583-4594, 2017.

Sun, S., He, M., Xu, M., Gao, S., Chen, Z., Zhang, X., Ruan, Z., Wu, X., Zhou, L., Liu, L., Lu, C., Guo, C., Liu, L., Yu, S., & Cai, X. "Bias-drift-free Mach-Zehnder modulators based on a heterogeneous silicon and lithium niobate platform", *Photonics Research*, vol. 8, no. 12, p. 1958-1963, 2020.

Sun, L., Xiao, J., Cai, Y., Shen, G., Liu, N., and Lu, C., "Complex-valued decision feedback equalizer for optical IMDD signals with adaptive manipulations in time and amplitude domains," *Optics Letters*, 47(17), 4391–4394, 2022.

Svarny, J., & Chladek, S., "Model-Based Bias Controller for a Mach-Zehnder Intensity Modulator", *Journal of Lightwave Technology*, 40(3), 720-727, 2022.

Tang, X., Qiao, Y., Chen, Y.-W., Lu, Y., and Chang, G.-K., "Digital Pre- and Post-Equalization for C-Band 112-Gb/s PAM4 Short-Reach Transport Systems," *Journal of Lightwave Technology*, vol. 38, no. 17, pp. 4683-4690, 2020.

Taniguchi, H., Yamamoto, S., Nakamura, M., Masuda, A., Kisaka, Y., Kanazawa, S., and Nakamura, H., "High-Capacity IM-DD Transmission in O-Band Using Advanced Maximum Likelihood Sequence Estimation Methods," *Journal of Lightwave Technology*, 41, 3372–3381, 2023.

Teng, J., Wang, S., Yin, Z., Chen, W., Fan-Yuan, G., Guo, G., Han, Z., & Guo, G., “Arbitrary bias control of LiNbO₃ based Mach-Zehnder intensity modulators for QKD system”, EPJ Quantum Technology, vol. 10, no. 33, 2023.

Tibaldi, A., Ghomashi, M., Bertazzi, F., and Ghione, G., “Organic Electro-Optic Mach-Zehnder Modulators: From Physics-Based to System-Level Modeling,” *physica status solidi (a)*, vol. 218(21), art. 2100390, 2021

Tibaldi, A., Ghomashi, M., Bertazzi, F., Goano, M., Vallone, M., & Ghione, G., “Plasmonic-organic hybrid electro/optic Mach-Zehnder modulators: from waveguide to multiphysics modal-FDTD modeling,” *Optics Express*, vol. 28, no. 20, pp. 29253-29271, 2020.

Tong, R., Dai, Z., Olsson, J., Huynen, I., Ruber, R., and Dancila, D., “Taper Transmission Line Based Measurement—A Thru-Only De-Embedding Approach,” *IEEE Transactions on Microwave Theory and Techniques*, 70(9), 4199 - 4210, 2022.

Torkaman, P., Latifi, S., Feng, K., and Yang, S., “Nonlinearity-Robust IM/DD THz Communication System via Two-Stage Deep Learning Equalizer”, *IEEE Communications Letters*, vol. 28, pp. 1805–1809, 2024.

Troost, C., Huber, R., Bell, A. R., van Delden, H., Filatova, T., Le, Q. B., Lippe, M., Niamir, L., Polhill, J. G., Sun, Z., and Berger, T., "How to keep it adequate: A protocol for ensuring validity in agent-based simulation models," *Environmental Modelling & Software*, vol. 159, 105559, 2023.

Truong, C., Oudre, L., Vayatis, N., “Selective review of offline change point detection methods”, *Signal Processing*, vol. 167, Art. 107299, 2020.

Tröls, M. A., Marchezan, L., Mashkoor, A., and Egyed, A., "Instant and global consistency checking during collaborative engineering," *Software and Systems Modeling*, vol. 21, pp. 2489-2515, 2022.

Tzimpragos, G., Kachris, C., Djordjevic, I. B., Cvijetic, M., Soudris, D., and Tomkos, I., “A Survey on FEC Codes for 100G and Beyond Optical Networks,” *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 209-221, 2016.

Unjhawala, H. M., Zhang, R., Hu, W., Wu, J., Serban, R., and Negrut, D., “Using a Bayesian-Inference Approach to Calibrating Models for Simulation in Robotics”, *Journal of Computational and Nonlinear Dynamics*, 2023.

van den Hout, M., van der Heide, S., and Okonkwo, C., “Digital Resolution Enhancer Employing Clipping for High-Speed Optical Transmission”, *Journal of Lightwave Technology*, vol. 38, no. 11, pp. 2897-2904, 2020.

Vanhoecke, M., Aimone, A., Argyris, N., Dris, S., Vaernewyck, R., Verheyen, K., Gruner, M., Fiol, G., Apostolopoulos, D., Avramopoulos, H., Torfs, G., Yin, X., and Bauwelinck, J., “Segmented Optical Transmitter Comprising a CMOS Driver Array and an InP IQ-MZM for Advanced Modulation Formats”, *Journal of Lightwave Technology*, vol. 35, no. 4, pp. 862-867, 2017.

Verbeek, S., Oomen, T., and Eielsen, A., “Glitch Compensation for a Digital-to-Analogue Converter”, *American Control Conference (ACC)*, pp. 751-757, 2021.

Verreault, A., Cicek, P., & Robichaud, A. “Oversampling ADC: A Review of Recent Design Trends,” *IEEE Access*, 12, 2024.

Vervoort, S., Saalberg, Y., and Wolff, M., “Mach-Zehnder Modulator Output in Time and Frequency Domain—Calculation and Experimental Confirmation,” *Photonics*, vol. 10, no. 3, Art. 337, 2023.

Vidal-Moreno, P., Becerril, C., Fernández-Ruiz, M., Martins, H., Martín-López, S., and González-Herráez, M., “Noise analysis in direct detection and coherent detection phase-sensitive optical time-domain reflectometry systems.”, *Optics Express*, vol. 31, no. 17, pp. 27450–27461, 2023.

Vilalta, R., Gifre, Ll., Casellas, R., Muñoz, R., Martínez, R., Mozo, A., “Applying Digital Twins to Optical Networks with Cloud-Native SDN Controllers”, *IEEE Communications Magazine*, pp. 129-133, 2023.

Villenas, V., Wu, D., Gültekin, N., Riani, M., and Alvarado, A., “On Geometric Shaping for 400 Gbps IM-DD Links with Laser Intensity Noise”, *Optical Fiber Communication Conference (OFC)*, Optica Publishing Group, paper Th1K.8, 2025. 2025.

Wang, J., Zhu, K., & Hossain, E. “Green Internet of Vehicles (IoV) in the 6G Era: Toward Sustainable Vehicular Communications and Networking,” *IEEE Transactions on Green Communications and Networking*, 6, 391-423, 2022.

Wang, S., Y. Liao, X. Cheng, and L. Zhang, “Role of Digital Twin in Optical Communication: Fault Management, Hardware Configuration, and Transmission Simulation,” *IEEE Communications Magazine*, vol. 59, no. 1, pp. 108-114, 2021.

Wang, Z., Gupta, R., Han, K., Wang, H., Ganlath, A., Ammar, N., & Tiwari, P. “Mobility digital twin: Concept, architecture, case study, and future challenges,” *IEEE Internet of Things Journal*, 9(18), 17452-17467, 2022.

Wang, D., Zhang, Z., Zhang, M., Fu, M., Li, J., Cai, S. “The Role of Digital Twin in Optical Communication: Fault Management, Hardware Configuration, and Transmission Simulation”, *IEEE Communications Magazine*, pp. 133-139, 2021.

Wang, D., Xu, Y., Li, J., Zhang, M., Li, J., Qin, J., “Comprehensive Eye Diagram Analysis: A Transfer Learning Approach”, *IEEE Photonics Journal*, vol. 11, no. 6, Art. no. 7905019, 2019.

Wang, D., Aziz, M., Helaoui, M., & Ghannouchi, F., “Augmented Real-Valued Time-Delay Neural Network for Compensation of Distortions and Impairments in Wireless Transmitters”. *IEEE Transactions on Neural Networks and Learning Systems*, 30(1), 242-254, 2019.

Wang, W., Li, F., Zou, D., Ni, W., Sui, Q., Yi, X., Cai, Y., and Li, Z., “Suppressed Noise Enhancement Equalization Combined With Nonlinear-MLSE Techniques for Bandwidth-Limitation IM/DD System”, *Journal of Lightwave Technology*, vol. 41, pp. 4655–4662, 2023.

Wang, X., & Kennedy, M., “Linearized Analysis and Quantization Error Minimization for Mid-Rise TDCs: A Tutorial”, *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 72, pp. 2986–2997, 2025.

Wang, F., Gao, R., Li, Z., Liu, J., Cui, Y., Xu, Q., Pan, X., Zhu, L., Guo, D., Chang, H., Zhou, S., Dong, Z., Zhang, Q., Tian, Q., Tian, F., Huang, X., Yan, J., Jiang, L., and Xin, X.,

“OAM mode-division multiplexing IM/DD transmission at 4.32 Tbit/s with a low-complexity adaptive-network-based fuzzy inference system nonlinear equalizer”, *Optics Letters*, vol. 49, no. 3, pp. 430–433, 2023.

Wang, K., Zhang, L., Yu, C., Wang, C., Chen, Y., Yu, J., “FPGA-based 4×29.4912 Gbit/s PS-PAM4 signal transmission with a low-complexity probabilistic shaping scheme,” *Optics Letters*, vol. 48(6), pp. 1514-1517, 2020.

Wang, K., Tang, R., Xiang, S., He, Y., He, Y., and Gui, X., “A 0.92-pJ/b 112-Gb/s PAM-4 Transmitter With Bandwidth and Linearity Enhanced Quasi-Voltage-Mode Driver and Reconfigurable Three-Tap T/2–T Variable Fractional-Spaced FFE in 28-nm CMOS,” *IEEE Transactions on Circuits and Systems I: Regular Papers*, 72, 2664–2675, 2025.

Wartak, M. S., "Computational Photonics: An Introduction with MATLAB", Cambridge University Press, Cambridge, 2013.

Wettlin, T., Calabrò, S., Rahman, T., Wei, J., Stojanović, N., Pachnicke, S., “DSP for High-Speed Short-Reach IM/DD Systems Using PAM,” *Journal of Lightwave Technology*, vol. 38(24), 6771-6778, 2020.

Widhianto, B., Tsui, H., and Chen, J., “Transition-Based Decision Feedback Equalization to Enhance Transmission in VCSEL-Based Short-Reach IM/DD System,” *IEEE Photonics Journal*, 17, 7201410, 2025.

Widrow, B., Kollár, I., and Liu, M.-C., “Statistical theory of quantization,” *IEEE Transactions on Instrumentation and Measurement*, vol. 45, no. 2, pp. 353-361, 1996.

Wiegart, T., Da Ros, F., Yankov, M., Steiner, F., Gaiarin, S., & Wesel, R., “Probabilistically Shaped 4-PAM for Short-Reach IM/DD Links With a Peak Power Constraint”, *Journal of Lightwave Technology*, 39(2), 400-405, 2020.

Williams, L., Sovacool, B., & Foxon, T. “The energy use implications of 5G: Reviewing whole network operational energy, embodied energy, and indirect effects,” *Renewable and Sustainable Energy Reviews*, 157, 112033, 2022.

Wilson, G., Aruliah, D. A., Brown, C. T., Chue Hong, N. P., Davis, M., Guy, R. T., Haddock, S. H. D., Huff, K. D., Mitchell, I. M., Plumbley, M. D., Waugh, B., White, E. P.,

and Wilson, P., “Best Practices for Scientific Computing”, PLOS Biology, vol. 12, no. 1, e1001745, 2014.

Witzens, J., Müller, J., and Moscoso-Martir, A., “Modification of Level Dependent ASE-Signal Beat Noise by Optical and Electrical Filtering in Optically Preamplified Direct Detection Receivers”, IEEE Photonics Journal, vol. 10, no. 1, 7500416, 2018.

Wu, J., Zhong, L., Dai, X., Yang, Q., Cheng, M., Deng, L., & Liu, D. “Numerical Analysis of Dither Induced Penalty for Automatic Bias Control Methods”, IEEE Photonics Journal, vol. 16, no. 1, 7100309, 2024.

Wu, T., Ho, C., & Chen, M. “A Flash-Based Non-Uniform Sampling ADC With Hybrid Quantization Enabling Digital Anti-Aliasing Filter,” IEEE Journal of Solid-State Circuits, 52, 2017.

Wu, K., Liga, G., Riani, J., and Alvarado, A., “Low-Complexity Soft-Decision Detection for Combating DFE Burst Errors in IM/DD Links”, Journal of Lightwave Technology, vol. 42, pp. 1395–1408, 2023.

Wu, Q., Xu, Z., Zhu, Y., Ji, H., Yang, Y., Qiao, G., Liu, L., Wang, S., Liang, J., Wei, J., He, Z., Zhuge, Q., Hu, W., “Gradient-descent noise whitening techniques for short-reach IM-DD optical interconnects with severe bandwidth limitation”, Optics Express, vol. 32, no. 2, pp. 1715–1727, 2024.

Wu, X., Karar, A., Zhong, K., Lau, A., and Lu, C., “Experimental demonstration of pre-electronic dispersion compensation in IM/DD systems using an iterative algorithm”, Optics Express, vol. 29, no. 16, pp. 24735–24749, 2021.

Wu, X., Zhang, J., Lau, A., Lu, C., “C-band 120-Gb/s PAM-4 transmissions over a 100-km dispersion-uncompensated SSMF using joint combined pulse shaping and low-complexity nonlinear equalization,” Optics Letters, vol. 47(19), pp. 5144-5147, 2022.

Xiang, M., Fu, S., Xu, O., Li, J., Peng, D., Gao, Z., Wang, Y., & Qin, Y., “Advanced DSP Enabled C-Band 112 Gbit/s/λ PAM-4 Transmissions With Severe Bandwidth-Constraint,” Journal of Lightwave Technology, vol. 40, no. 4, pp.987-996, 2022.

Xiao, J.-L., and Huang, Y.-Z., “Numerical analysis of gain saturation and noise figure in inhomogeneously broadened quantum-dot semiconductor optical amplifiers”, *IEEE Journal of Quantum Electronics*, vol. 44, no. 5, pp. 448 - 455, 2008.

Xie, Y., Wang, Y., Kandeepan, S., & Wang, K., "Machine learning applications for short reach optical fiber communications", *Photonics*, vol. 9, no. 1, p. 30, 2022.

Xie, Y., Kandeepan, S., and Wang, K., “An Analytical Model of OSNR and SNR for Short Reach IM-DD Systems with MZM”, *2022 Asia Communications and Photonics Conference (ACP)*, pp. 1826–1830, 2022.

Xie, Y., He, P., Li, W., and Li, N., “A Decision Feedback Equalization Algorithm Based on Simplified Volterra Structure for PAM4 IM-DD Optical Communication Systems”, *Applied Sciences*, 2023.

Xin, H., Zhang, K., Kong, D., Zhuge, Q., Fu, Y., Jia, S., Hu, W., and Hu, H., “Nonlinear Tomlinson-Harashima precoding for direct-detected double sideband PAM-4 transmission without dispersion compensation”, *Optics Express*, vol. 27, no. 14, pp. 19156-19167, 2019.

Xu, Z., Ji, T., Wu, Q., Lu, W., Ji, H., Yang, Y., Qiao, G., Tang, J., Cheng, C., Liu, L., Wang, S., Liang, J., Sun, Z., Fan, L., Wei, J., and Shieh, W., "Advanced Neural Network-Based Equalization in Intensity-Modulated Direct-Detection Optical Systems: Current Status and Future Trends," *Photonics*, vol. 11, no. 8, article 702, 2024.

Xu, H., Lai, M., Zheng, X., Wang, Z., Xu, J., Li, S., Lv, F., Liu, M., Li, W., Wen, Z., Guo, X., Wang, X., Jin, Z., and Liu, X., “A 112-Gb/s PAM-4 Retimer Transceiver With Jitter-Filtering Clocking Scheme and BER Optimization Technique in 28-nm CMOS”, *IEEE Journal of Solid-State Circuits*, vol. 60, pp. 2305–2318, 2025.

Xu, T., Li, Z., Peng, J., Tan, A., Song, Y., Li, Y., Chen, J., and Wang, M., “Decoding of 10-G Optics-Based 50-Gb/s PAM-4 Signal Using Simplified MLSE,” *IEEE Photonics Journal*, 10, 7905508, 2018.

Xue, L., Yi, L., Lin, R., Huang, L., and Chen, J., “SOA pattern effect mitigation by neural network based pre-equalizer for 50G PON”, *Optics Express*, vol. 29, no. 16, pp. 24714–24722, 2021.

Yan, A., Li, G., Xing, S., Shen, W., Hu, Y., Sun, A., Shi, J., Li, Z., Shen, C., Chi, N., Zhang, J., “Net-432-Gb/s/ λ PS-PAM transmission based on advanced DSP and single DAC for 400 G/ λ IM/DD links,” *Optics Letters*, Vol. 49, Issue 7, pp. 1656-1659, 2024.

Yan, A., Xing, S., Li, Z., Li, G., Sun, A., Luo, P., Shi, J., Zhang, H., Wang, D., Xiao, X., He, Z., Chi, N., & Zhang, J., “Adaptive Partial-Response Neural Network Equalization for Bandwidth-Limited PAM Transmission in Intra-Datacenter Interconnect,” *Journal of Lightwave Technology*, vol. 42, pp. 2762–2773, 2024,

Yang, C., Liang, P., Fu, L., Cui, G., Huang, F., Teng, F., & Bangash, Y. “Using 5G in Smart Cities: A Systematic Mapping Study”. *Intelligent Systems with Applications*, 14, 200065, 2022.

Yang, F., Gu, S., “Industry 4.0: A revolution that requires technology and national strategies,” *Complex & Intelligent Systems*, Volume 7, pages 1311-1325, 2021.

Yang, H., Niu, Z., Zhao, H., Xiao, S., Hu, W., Yi, L., "Fast and Accurate Waveform Modeling of Long-Haul Multi-Channel Optical Fiber Transmission Using a Hybrid Model-Data Driven Scheme," *Journal of Lightwave Technology*, vol. 40, no. 14, pp. 4571-4580, 2022.

Yang, Y., Li, D., Qi, N., and Wang, B., “A 2 pA/ $\sqrt{\text{Hz}}$ Input-Referred Noise TIA in 180-nm CMOS With 2.5 GHz Bandwidth for Optical Receiver”, *IEEE Solid-State Circuits Letters*, vol. 8, pp. 189–192, 2025.

Ye, T., Su, X., Yang, C., Li, J., Tao, Z., Nakashima, H. (2022). “Noise analysis for communication systems using high-speed DAC and ADC”. *ECOC 2022*.

Yi, H., Chang, F., Tan, P., Huang, B., Wu, Y., Xu, Z., et al. “Design of infinite impulse response maximally flat stable digital filter with low group delay,” *Scientific Reports*, 15, 2025.

Yigitcanlar, T., Kamruzzaman, M., Buys, L., Ioppolo, G., Sabatini-Marques, J., da Costa, E. M., & Yun, J. J. “Understanding ‘Smart Cities’: Intertwining Development Drivers with Desired Outcomes in a Multidimensional Framework.” *Cities*, 81, 145-160, 2018.

Yin, M., Zou, D., Wang, W., Li, F., and Li, Z., “Transmission of a 56 Gbit/s PAM-4 signal with low-resolution DAC and pre-equalization only over 80 km fiber in C-band IM/DD systems for optical interconnects,” *Optics Letters*, vol. 46, no. 22, pp. 5615-5618, 2021.

Yoshida, T., Alvarado, A., Karlsson, M., & Agrell, E., “Post-FEC BER Benchmarking for Bit-Interleaved Coded Modulation With Probabilistic Shaping,” *Journal of Lightwave Technology*, vol. 38, no. 16, pp. 4292–4306, 2020.

Yoshida, T., Karlsson, M., & Agrell, E., “Performance Metrics for Systems With Soft-Decision FEC and Probabilistic Shaping,” *IEEE Photonics Technology Letters*, vol. 29, pp. 2111–2114, 2017.

Yu, H., Michalka, T., Larbi, M., & Swaminathan, M. “Behavioral Modeling of Tunable I/O Drivers With Preemphasis Including Power Supply Noise”. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 28, 233-242, 2020.

Yu, Y., Che, Y., Bo, T., Kim, D., and Kim, H., “Reduced-state MLSE for an IM/DD system using PAM modulation”, *Optics Express*, vol. 28, no. 26, pp. 38505–38515, 2020.

Yu, Y., Choi, M., Bo, T., He, Z., Che, Y., and Kim, H., “Low-Complexity Second-Order Volterra Equalizer for DML-Based IM/DD Transmission System”, *Journal of Lightwave Technology*, vol. 38, pp. 1735–1746, 2020.

Yuan, A., and Savory, S., “Data-Driven Modeling of a Direct Detection System With an Electro-Absorption Modulated Laser”, *Journal of Lightwave Technology*, vol. 43, pp. 5545–5557, 2025.

Yuan, F., Al-Tae, A., Ye, A., and Sadr, S., “Design techniques for decision feedback equalisation of multi-giga-bit-per-second serial data links: a state-of-the-art review,” *IET Circuits, Devices & Systems*, 8, 118–130, 2014.

Zeng, F., Pang, C., & Tang, H. “Sensors on Internet of Things systems for the sustainable development of smart cities: A systematic literature review,” *Sensors*, 24(7), 2074, 2024.

Zeng, W., Wang, Y., and Wang, Z., “A de-embedding method based on combining time and frequency domains,” *Scientific Reports*, 15, 18637, 2025.

Zhang, C., Chen, F., Wang, L., Wang, L., & Patrick Yue, C. “Recent Advances of High-Speed Short-Reach Optical Interconnects for Data Centers.” *IEEE Open Journal of the Solid-State Circuits Society*, 5, 86-100, 2025.

Zhang, J., Yu, J., Wey, J.S., Li, X., Zhao, L., Wang, K., Kong, M., Zhou, W., Xiao, J., Xin, X., & Zhao, F., "SOA Pre-Amplified 100 Gb/s/ λ PAM-4 TDM-PON Downstream Transmission Using 10 Gbps O-Band Transmitters", *Journal of Lightwave Technology*, vol. 38, pp. 185-193, 2020.

Zhang, L., Chen, L., and Bao, X., “Unveiling Delay-Time-Resolved Phase Noise Statistics of Narrow-Linewidth Laser via Coherent Optical Time Domain Reflectometry”, *Optics Express*, vol. 28, no. 5, pp. 6719-6733, 2020. DOI: 10.1364/OE.387185.

Zhang, Z., and Yariv, A., “A General Relation Between Frequency Noise and Lineshape of Laser Light”, *IEEE Journal of Quantum Electronics*, early access, 56(3), 1100105, 2020.

Zhang, J., Lin, Z., Wu, X., Liu, J., Lau, A., Guo, C., Lu, C., and Yu, S., “Low-complexity sparse absolute-term based nonlinear equalizer for C-band IM/DD systems”, *Optics Express*, vol. 29, no. 14, pp. 21891–21901, 2021.

Zhang, J., Tan, H., Hong, X., Liu, J., Guo, C., Fei, C., Wu, X., Lau, A., Yu, S., and Lu, C., “Comparison of low-complexity sparse and weight-sharing nonlinear equalizers for C-band 100-Gbit/s DSB PAM-4 transmission over 60-km SSMF”, *Optics Express*, vol. 30, no. 20, pp. 20, pp.

Zhang, J., Tan, H., Lau, A., Li, Z., and Lu, C., “Low-complexity cluster-assisting look-up-table-based Volterra decision-feedback equalizer for IM/DD systems”, *Optics Letters*, vol. 50, no. 3, pp. 992–995, 2025.

Zhang, T., Liu, W., Zhong, Z., Li, H., Gong, Y., He, Q., Luo, M., Liu, S., “A TCN Based Data-Driven Method Combined With TL for Modeling in Experimental IM-DD System,” *IEEE Photonics Technology Letters*, vol. 37(2), pp. 105 – 108, 2025.

Zhang, X., Huo, J., Bai, H., Qin, P., Wei, H., and Long, K., “Performance Investigation of Joint LUT and GS Algorithm at the Transceiver for Nonlinear and CD Compensation”, *Photonics*, 11(7), 665, 2024.

Zhao, X., Zhang, J., Zhou, J., Hu, S., Xu, B., Zhang, Q., and Qiu, K., “Collaborative FFE-assisted simplified soft-output MLSE with LDPC for high-speed IM-DD transmissions”, *Optics Express*, vol. 32, no. 3, pp. 3866–3873, 2024.

Zheng, X., Ding, H., Zhao, F., Wu, D., Zhou, L., Wu, J., Lv, F., Wang, J., and Liu, X., “A 50–112-Gb/s PAM-4 Transmitter With a Fractional-Spaced FFE in 65-nm CMOS,” *IEEE Journal of Solid-State Circuits*, 55, 1864–1876, 2020.

Zhong, K., Zhou, X., Huo, J., Yu, C., Lu, C., Lau, A. P. T., “Digital Signal Processing for Short-Reach Optical Communications: A Review of Current Technologies and Future Trends”, *Journal of Lightwave Technology*, 36(2), 377-400, 2018.

Zhou, H., Li, Y., Liu, Y., Yue, L., Gao, C., Li, W., and Qiu, J., “Recent Advances in Equalization Technologies for Short-Reach Optical Links Based on PAM4 Modulation: A Review”, *Applied Sciences*, vol. 9, no. 11, 2019.

Zhou, X., Urata, R., and Liu, H., "Beyond 1 Tb/s Intra-Data Center Interconnect Technology: IM-DD OR Coherent?", *Journal of Lightwave Technology*, vol. 38, pp. 475-484 , 2020.

Zhou, Z., Odedeyi, T., Kelly, B., O'Carroll, J., Phelan, R., Darwazeh, I., & Liu, Z., “Impact of Analog and Digital Pre-Emphasis on the Signal-to-Noise Ratio of Bandwidth-Limited Optical Transceivers”, *IEEE Photonics Journal*, vol. 12, pp. 1–12, 2020.

Zhou, J., Zhang, J., Zhao, X., Jin, T., Hu, S., Lin, H., Zhenming, Y., Zhang, Q., Xu, B., and Qiu, K., “Advanced nonlinearity equalizer with TC-NL-MLSE for transmitting beyond 200-Gb/s PAM-8 in IM/DD systems,” *Optics Express*, 30(21), 37416–37425, 2022.

Zhou, J., Zhang, J., Zhao, X., Wang, R., Li, J., Hu, S., Zhang, Q., Yang, Q., Xu, B., and Qiu, K., “MLSE-assisted decision feedback equalizer for error-propagation suppression in high-speed IM-DD transmission systems,” *Chinese Optics Letters*, 2024.

Zhou, L., Fang, X., Zhu, Y., Zhu, J., Li, Y., Chen, Z., Nie, X., Yang, C., Zhou, X., He, Z., Wang, L., Li, K., Zhang, F., “Up to 182 GBd PAM-4 optical interconnects eliminating receiver-side equalization DSP,” *Optics Letters*, Vol. 50(7), pp. 2263-2266, 2025.

Zhu, P., Yoshida, Y., Akahane, K., and Kitayama, K., "High-speed reach-extended IM-DD system with low-complexity DSP for 6G fronthaul [Invited]", *Journal of Optical Communications and Networking*, vol. 16, pp. A24-A32, 2023.

Zhu, Y., Miao, X., Wu, Q., Yin, L., & Hu, W., “Imbalanced Mach-Zehnder Modulator for Fading Suppression in Dispersion-Uncompensated Direct Detection System”, *Electronics*, 10(22), 2866, 2021.

Zhu, Y., Diaz, J., Kaile, S., Yi, I., Liu, T., Hoyos, S., & Palermo, S., “A Jitter-Robust 40 Gb/s ADC-Based Multicarrier Receiver Front-End With 4-GS/s Baseband Pipeline-SAR ADCs in 22-nm FinFET”, *IEEE Journal of Solid-State Circuits*, vol. 58, pp. 662–676, 2023.

Zhu, P., Yoshida, Y., Kanno, A., and Kitayama, K., “Analysis and Demonstration of Low-Complexity Joint Optical-Electrical Feedforward Equalization (OE-FFE) For Dispersion-Limited High-Speed IM/DD Transmission”, *Journal of Lightwave Technology*, vol. 41, pp. 477–488, 2023.

Zhuge, Q., et al., “Building a Digital Twin for Intelligent Optical Networks [Invited Tutorial],” *Journal of Optical Communications and Networking*, vol. 15, no. 8, pp. C242-C262, 2023.

Zibar, D., Piels, M., Jones, R., and Schaeffer, C.G., “Machine Learning Techniques in Optical Communication,” *Journal of Lightwave Technology*, vol. 34, pp. 1442-1452, 2016.

Ó Dúill, S. P., Landais, P., and Barry, L. P., “Estimation of the Performance Improvement of Pre-Amplified PAM4 Systems When Using Multi-Section Semiconductor Optical Amplifiers”, *Applied Sciences*, vol. 7, no. 9, art. 908, 2017.



Università telematica delle
Camere di Commercio Italiane

The PhD scholarship co-funded with resources from the European Union – *NextGeneration EU*
National Recovery and Resilience Plan (PNRR), Mission 4, Component 2 “*From Research to Business*” –
Investment 3.3 “*Introduction of innovative PhD programmes that meet the innovation needs of enterprises and
promote the recruitment of researchers by companies*” – CUP D83C22001880003



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Università telematica delle
Camere di Commercio Italiane