

Lumped Silicon Photonic Mach-Zehnder Modulators for High-Speed Optical Interconnects

Brian Murray

BSc

110302655

**Thesis submitted for the degree of
Doctor of Philosophy**



NATIONAL UNIVERSITY OF IRELAND, CORK

COLLEGE OF SCIENCE, ENGINEERING AND FOOD SCIENCE

DEPARTMENT OF PHYSICS

TYNDALL NATIONAL INSTITUTE

PHOTONICS SYSTEMS GROUP

7 August 2022

Head of Department: Prof John McInerney

Supervisors: Prof Paul Townsend

Dr Cleitus Antony

Dr Giuseppe Talli

Contents

List of Figures	iv
List of Tables	xi
Acknowledgements	xiii
Abstract	xv
1 Photonic Systems and Data Centres	1
1.1 Introduction	1
1.2 Context and Motivation	1
1.2.1 Data Centres: from Copper to Single-Mode Fibre	6
1.2.2 Silicon Photonics	9
1.3 Multilevel Modulation Formats	13
1.3.1 Four-Level Pulse Amplitude Modulation	13
1.3.2 Electrical Duobinary Modulation	14
1.3.3 Bit Error Rate	17
1.4 Summary	22
2 Fundamentals of Silicon Photonic Mach-Zehnder Modulators	23
2.1 Introduction	23
2.2 The Mach-Zehnder Modulator	23
2.2.1 The Mach-Zehnder Interferometer Structure	23
2.2.2 Silicon Waveguides	27
2.2.3 Optical Fibre to Silicon Waveguide Coupling	29
2.2.4 Modulation with an MZM	30
2.2.5 MZM Drive Formats	32
2.2.6 Unbalanced MZMs	34
2.2.7 Thermo-optic Effect in Silicon	36
2.3 Modulation in Silicon	39
2.3.1 Complex Refractive Index and Kramers-Kronig	39
2.3.2 Achieving Phase Shift in Silicon	40
2.3.3 Silicon Phase Shifter Structures	44
2.4 Lumped vs. Travelling-Wave MZMs	45
2.5 Power Consumption and Drive Voltage	51
2.6 State of the Art	55
2.7 Summary	56
3 Optimisation of Lumped SiP MZMs	57
3.1 Introduction	57
3.2 Characterisation of SiP MZMs	58
3.2.1 Voltage-Current Characteristics	58
3.2.2 Measuring Phase Shift and Absorption with an OSA	59
3.2.3 Measuring Bandwidth with an AWG and Oscilloscope	62
3.3 The Roll of Doping Levels in SiP MZMs	66
3.3.1 Modelling the phase shift and absorption of SiP MZMs	66
3.3.2 Influence of Doping on Extinction Ratio, Insertion Loss and Level Mismatch Ratio	71

3.3.3	Modelling the RC Bandwidth of SiP MZMs	76
3.4	Switched Capacitor Drive Scheme	83
3.4.1	Switched-Capacitor Drive Scheme for PAM4	83
3.4.2	Optimisation of Bias and Doping Levels for Switched-Capacitor Lumped MZMs	86
3.5	Full Silicon Photonic Link Demonstration	92
3.5.1	Experimental Demonstration	92
3.5.2	Operation of the SiP Link without Optical Amplification	95
3.6	Summary	99
4	Predistortion for High-Speed Lumped SiP MZMs	101
4.1	Introduction	101
4.2	Experimental Evaluation of Precompensation for Lumped MZMs	102
4.2.1	Travelling-Wave and Lumped MZM Frequency Response Comparison	102
4.2.2	Experiment Setup	104
4.2.3	Results	105
4.3	Designing Precompensation and Modelling Power Consumption	111
4.3.1	Voltage and Current Behaviour in an RC circuit	112
4.3.2	Precompensation for an RC circuit	114
4.3.3	Determining a Matched Precompensation Filter for Arbitrary Waveforms	117
4.4	Designing Optimal Precompensation for the Lumped MZM	119
4.5	Modelling the Frequency Response of the Lumped MZM	125
4.5.1	Deriving an Accurate Equivalent Circuit Model of a Lumped MZM from S11 Measurements	125
4.5.2	Deriving the S21 Frequency Response of the Lumped MZM from the Equivalent Circuit Model	131
4.5.3	Effect of Optical Transit Time on Lumped MZM Frequency Response	133
4.5.4	Effect of Removing the Termination Resistor	137
4.6	Improving Lumped MZM Bandwidth through Driver Considerations	143
4.6.1	Method 1: Use Predistortion	143
4.6.2	Method 2: Reduce Source Resistance	147
4.6.3	Method 3: Predistortion & Reduced Source Resistance	154
4.6.4	Summary of Bandwidth-Boosting Drive Methods	157
4.7	Summary	160
5	Conclusions	162
5.1	Contributions Overview	162
5.2	Future Work	165
A	BER of an EDB Signal	167
B	Sinusoidal Current and Voltage in an RC Circuit	170
C	Thévenin Equivalent Circuit Derivation for Lumped MZM	173

D A Brief Exhibition	177
List of Acronyms and Initialisms	180
List of Publications	183
Published Journals and Conference Contributions	183
Unpublished Conference Contributions	183

List of Figures

1.2	Number of data centres in countries in the European region as of 2019. Inset: breakdown of Irish data centre categories into private, colocation and hyperscale as a percentage of total power capacity.	3
1.3	A typical data centre architecture as described by Facebook. . .	5
1.5	Attenuation profile and transmission windows for optical fibre. .	8
1.6	Progression of Ethernet standards in terms of (a) bit rate, (b) distance, and (c) bit-rate–distance product.	10
1.7	Simulated waveforms and corresponding eye diagrams for (a) NRZ, (b) PAM4 and (c) EDB.	15
1.9	Normal distribution of binary symbol samples and the receiver decision regions.	18
1.10	Comparison of simulated BER curves for NRZ, EDB and PAM4. .	21
2.1	Basic structure of an MZM, with phase shifters in both arms. . .	24
2.2	(a) Output electric field of an MZM, in terms of magnitude (abs) and phase (arg), against net phase variation in the arms. (b) Resulting output power of the MZM.	25
2.3	Required change in effective refractive index at 1550 nm for phase shift with various MZM lengths.	26
2.4	Simulated intensity profile of the optical TE mode at 1550 nm in (a) SiP strip waveguide, and (b) SiP rib waveguide.	28
2.5	(a) Scale cross-section comparison of single-mode fibre core and typical silicon waveguide. (b) Not-to-scale profile of a 1D silicon grating coupler as an out-coupling device from a silicon waveguide.	30
2.6	(a) 10 GHz sinusoidal drive voltage. (b) Corresponding MZM output power against time.	32
2.7	Simplified MZM driving circuitry showing (a) single-drive and (b) dual-drive schemes.	33
2.8	(a) MZM FSR variation against length imbalance in the C-band and O-band. (b) MZM transmission spectrum for various length imbalances.	35
2.9	(a) Variation of the transmission spectrum of an unbalanced MZM with temperature. (b) Variation in transmission with temperature change for various length imbalances.	37
2.10	(a) Balanced MZM transmission spectrum (b) Balanced MZM power transmission against arm temperature difference at 1550 nm.	38
2.11	(a) Refractive index changes and (b) absorption coefficient changes due to concentrations of electrons and holes. (c) Total refractive index change from both electrons and holes. (d) Resulting optical power loss in dB/mm.	43

2.12	Cross sections of silicon phase shifters made from rib waveguides using (a) carrier accumulation design, (b) carrier injection design, and (c) carrier depletion design.	44
2.13	Measured transmission spectrum dependence on voltage for the SiP MZM.	48
2.14	Micrographs of (a) dual-drive SiP Lumped MZM, and (b) dual-drive SiP TW-MZM.	50
2.15	(a) Transmission line diagram of a TW-MZM with $50\ \Omega$ driver and matching termination resistor. (b) Simple equivalent circuit diagram for the same structure.	51
2.16	Equivalent circuit of a lumped MZM with $50\ \Omega$ driver.	52
2.17	(a) Energy per bit for a single MZM arm for the two drive scenarios. (b) The ratio of the TW-MZM energy consumption against that of the lumped MZM.	54
3.1	VI curve for one arm of the lumped MZM, showing diode behaviour.	58
3.2	Setup for measuring phase shift and absorption characteristics of SiP MZM against voltage.	59
3.3	Measured transmission spectrum dependence on voltage for the SiP MZM.	60
3.4	(a) Measured phase shift and (b) change in absorption against bias voltage, for a 1.5 mm SiP depletion-mode phase shifter. . .	61
3.5	Typical optical testbed for systems tests of an MZM, suitable also for frequency response measurements using SPHS.	62
3.6	Simulated input and output for flat-spectrum SPHS passed through an ideal fourth-order, 12.5 GHz Bessel filter, shown as: (a) time domain waveform, (b) amplitude-frequency response, and (c) phase-frequency response.	64
3.7	Frequency response of an unterminated SiP MZM at various bias voltages, measured using SPHS.	65
3.8	3D diagram of a silicon p-n junction phase shifter section (not to scale).	66
3.9	Flow chart depicting the steps used to obtain the simulated phase shift and absorption characteristics of the silicon depletion-mode phase shifter.	68
3.10	Depletion width shown according to the p-type region, n-type region and total.	69
3.11	Propagation loss due to free-carrier absorption, plotted against bias voltage for various doping levels in an MZM with 1.5 mm active length.	72
3.12	Phase shift in a single MZM arm, plotted against bias voltage for various doping levels in an MZM with 1.5 mm active length. The phase shift is shown relative to a reference bias point of -4 V . .	73
3.13	Extinction ratio against total peak-to-peak RF voltage in a dual-drive push-pull 1.5 mm MZM operating a quadrature, with applied bias of -4 V , for various doping levels.	74

3.14	Level mismatch ratio against total peak-to-peak RF voltage in a dual-drive push-pull 1.5 mm MZM operating a quadrature, with applied bias of -4 V , for various doping levels.	75
3.15	(a) Carrier mobility and (b) material resistivity against carrier concentration for silicon doped with boron (p-type) and phosphorus (n-type).	78
3.16	(a) Simulated capacitance, (b) resistance and (c) intrinsic RC 3 dB bandwidth of a silicon phase shifter.	79
3.17	(a) Simulated phase shifter resistance and (b) intrinsic 3 dB bandwidth against central junction doping, with intermediate doping of $2 \times 10^{18}\text{ cm}^{-3}$. (c) The intrinsic bandwidth over a wider range of junction doping, showing the effect of intermediate doping levels.	80
3.18	(a) 3 dB RC bandwidth for various lengths of MZMs, with intermediate doping of $2 \times 10^{18}\text{ cm}^{-3}$, with and without a $50\ \Omega$ driver. (b) The bandwidth for the case of the 1.5 mm MZM with a $50\ \Omega$ driver, for various intermediate dopings.	81
3.19	RC 3 dB bandwidth against driver output impedance, for MZMs with intermediate and central junction doping fixed at $2 \times 10^{18}\text{ cm}^{-3}$ and $2 \times 10^{17}\text{ cm}^{-3}$, respectively, for various active region lengths	82
3.20	(a) Typical $50\ \Omega$ drive scheme for a silicon phase shifter. (b) Capacitive drive scheme, where the driver output impedance is reduced and a series capacitor is added to the driver.. . . .	84
3.21	RC frequency responses for three different lumped drive schemes: $50\ \Omega$, $5\ \Omega$, and $5\ \Omega$ with series capacitor.	84
3.22	Switched-capacitor, dual-drive, push-pull, PAM4 drive scheme for the silicon MZM.	85
3.23	Cross section of the SiP rib-waveguide phase shifter, showing the optical mode, the p-type and n-type doping regions, and the depletion region.	87
3.24	Phase shift and absorption against bias voltage for various levels of doping in a 1 mm SiP phase shifter.	88
3.25	Half-V _{pi} and insertion loss against bias for various levels of doping in a silicon MZM with 1.5 mm phase shifters.	88
3.26	Optical modulation amplitude against bias voltage for various doping levels in a silicon MZM with 1.5 mm phase shifters. The applied peak-to-peak RF voltage is 1 V	89
3.27	Level mismatch ratio in PAM4 and PAM8 signals, plotted against bias voltage for a silicon MZM	90
3.29	Full SiP link, including switched-capacitor PAM4 lumped MZM Tx, and silicon photoreceiver Rx.	92
3.30	Measured BER curves for SiP PAM4 link, for various lengths of SSF, with (a) no equalisation applied and (b) 14-tap fractionally spaced FFE applied.	94

3.31 (a) ER and (b) OMA against thermal or wavelength-tuned phase offset from quadrature in the MZM.	96
3.32 (a) ER and (b) OMA against bias voltage in the two MZM arms.	97
3.33 Rx sensitivity for the simulated 15 Gbaud SiP switched-capacitor PAM4 link at quadrature.	99
4.1 Normalised electro-optic frequency response of the terminated TW-MZM, unterminated TW-MZM, and lumped MZM.	103
4.2 Experimental setup for 25 Gb/s and 50 Gb/s modulation with a terminated (setup A, 50 Ω load attached) and unterminated (setup B, 50 Ω load removed) SiP TW-MZM, and for a lumped SiP MZM (setup C).	103
4.3 Effect of the FIR predistortion filter on (a) the unterminated MZM and (b) the lumped MZM.	106
4.4 BER comparison of the unterminated MZM with no predistortion applied, terminated MZM with the first-order FIR predistortion applied, and lumped MZM with same predistortion applied for 25 Gb/s NRZ modulation.	106
4.5 (a) 25 Gb/s terminated MZM NRZ eye. (b) 25 Gb/s unterminated MZM NRZ eye with first-order FIR Tx filter. (c) 25 Gb/s lumped MZM NRZ eye with first-order FIR Tx filter.	107
4.6 (a) BER performance of the unterminated MZM for 25 Gb/s NRZ, EDB and PAM4. (b) 25 Gb/s PAM4 eye diagram. (c) 25 Gb/s EDB eye diagram. The corresponding BER curves, PAM4 and EDB eye diagrams for the lumped MZM are shown in (d), (e) and (f) respectively.	108
4.7 Eye diagrams obtained for the B2B 50Gb/s PAM4 and EDB signals using a 16-tap fractionally spaced FFE at the receiver.	109
4.8 (a) The unterminated MZM's BER performance for 50 Gb/s EDB and PAM4 with fractionally spaced FFE over B2B and 10 km of fibre. Also shown is symbol spaced FFE for EDB over 10 km of fibre. (b) The corresponding curves for the lumped MZM, showing error flooring.	111
4.9 Flow chart describing the stages of the circuit model.	113
4.10 (a) Frequency response of the resistor and capacitor in the RC circuit. (b) Total power dissipated by an applied sinusoidal voltage. (c) Energy dissipated per cycle of the sinusoid. (d) Energy dissipated per bit for an applied PRBS waveform.	114
4.11 (a) Energy dissipated per bit in the RC circuit for various bandwidth boost factors shown against NRZ symbol rate. (b) A focus on energy dissipated for 25 Gbaud NRZ with inset corresponding eye diagrams. (c) Sensitivity improvement for 25 Gbaud NRZ against boost factor. (d) Corresponding sensitivity improvement for 25 Gbaud PAM4.	116
4.12 Measured (a) frequency response and (b) impulse response of the lumped MZM, with and without smoothing.	120

4.13	Comparison of the matched precompensation filter for 25 Gbaud PAM4 in terms of (a) frequency response and (b) impulse response, with and without the additional taps to compensate for the RF reflections in the drive cables.	121
4.14	Measured 50 Gb/s eye diagrams for (a) PAM4 and (b) EDB waveforms obtained from the lumped MZM with matched precompensation filter.	122
4.16	Equivalent circuit of the driver and phase shifter for a lumped MZM, including parasitics.	125
4.17	(a) Smith chart plot of the measured S11 of a reference unloaded transmission line after calibration. (b) Comparison of the measured S11 of the lumped MZM under forward bias (1 V) and reverse bias (−1 V). All measurements run from 10 MHz to 40 GHz.	127
4.18	Measured S11 parameter of the lumped MZM in terms of (a) magnitude and (b) phase.	128
4.19	Training curves for the final fit of the equivalent circuit model to the measured S11 parameters for various bias voltages.	130
4.20	Fit of the equivalent circuit model to the measured S11 parameter at various bias voltages in terms of (a) its real part and (b) imaginary part.	131
4.21	Conversion of the lumped MZM's equivalent circuit model from its standard form to its Thévenin equivalent.	132
4.22	Similarity of optical transit time effect and fourth-order Bessel response. $L_{mzm} = 1.5$ mm.	135
4.23	Comparison of the measured S21 frequency response of the lumped MZM with the model (with and without the optical transit time effect).	136
4.24	Comparison of the frequency response in terms of (a) amplitude squared and (b) group delay vs frequency for the LRC model with and without the 50 Ω termination.	139
4.25	Simulated PAM4 eye diagrams at 10 Gbaud and 25 Gbaud for the terminated and unterminated MZM.	140
4.26	Simulated PAM4 BER curves at 10 Gbaud and 25 Gbaud for the terminated and unterminated MZM, with and without the effects of the optical transit time.	140
4.27	Comparison of the terminated and unterminated MZM in terms of (a) energy dissipated for sinusoidal waveforms and (b) energy dissipated per bit for NRZ waveforms. Also compared for the MZMs are the waveforms for (c) source voltage and (d) drawn current in the case of 10 Gbaud NRZ.	142
4.28	The effect of digital precompensation on the frequency response of the unterminated MZM for (a) amplitude and (b) group delay. The response of the terminated MZM is shown for comparison.	144

4.29	Comparison of the simulated 25 Gbaud PAM4 eye diagrams produced by the various frequency responses in Figure 4.28. The eyes are respectively from the (a) terminated MZM, (b) unterminated MZM, and (c) the unterminated MZM with precompensation.	144
4.30	Energy dissipated per cycle for sinusoidal drive waveforms in the circuit in the case of the terminated MZM, unterminated MZM, and precompensated unterminated MZM.	145
4.31	Comparison of the (a) source voltage, (b) drawn current and (c) load voltage waveforms for 25 Gbaud NRZ for the terminated, unterminated and precompensated-unterminated MZMs. . . .	146
4.32	Effect of lowering source resistance on the frequency response of the unterminated MZM in terms of (a) magnitude and (b) group delay.	148
4.33	(a) Comparison of the 3 dB bandwidths as the source resistance is lowered for the LRC model and the simple RC model. (b) Comparison of the energy dissipated in the circuits for 25 Gbaud NRZ and PAM4 signals as the source resistance is lowered. . . .	149
4.34	Required Rx power for a BER of 1e-3 for 25 Gbaud NRZ and PAM4 signals when sent by the unterminated MZM. The equivalent for a simple RC circuit is shown for comparison.	150
4.35	Effect of lowering source resistance on the frequency response of the unterminated MZM with high inductance in terms of (a) magnitude and (b) group delay.	152
4.36	(a) Comparison of the 3 dB bandwidths as the source resistance is lowered for the lumped MZM with low inductance and high inductance. (b) Comparison of the energy dissipated in the circuits as the source resistance is lowered.	153
4.37	Required Rx power for a BER of 1e-3 for 25 Gbaud NRZ and PAM4 signals from the lumped MZM, plotted against source resistance, comparing the cases of low and high inductance.	154
4.38	Required Rx power for a BER of 1e-3 for 25 Gbaud NRZ and PAM4 signals when sent by the unterminated MZM without and with precompensation.	156
4.39	(a) Peak-to-peak voltage required for a 25 Gbaud NRZ signal with 4 dbER as source resistance is reduced for the unterminated MZM for the cases of without and with precompensation. (b) Corresponding energy dissipated in the circuit, shown for NRZ and PAM4.	157
4.40	(a) Current drawn for unterminated MZM with precompensation in the cases of 50 Ω (poor bandwidth regime) and 20 Ω (inductive peaking regime) source resistance. (b) The precompensated source voltages for the two cases, showing how the poor bandwidth regime requires more voltage.	158

4.41	Summary of the discussed lumped MZM drive options for 25 Gbaud PAM4 in terms of required drive voltage to achieve a 4 dB ER, also showing the energy dissipated per bit and eye quality.	159
A.1	Normal distribution of duobinary symbol samples and the receiver decision regions.	167
B.1	A simple RC circuit diagram, which is often used to model the frequency behaviour of lumped MZMs.	170
B.2	RC frequency response ($R = 50 \Omega$, $C = 500 \text{ fF}$).	172
C.1	The lumped MZM's equivalent circuit model and its Thévenin equivalent.	173
C.2	Process for obtaining the Thévenin equivalent circuit used in Chapter 4 of this thesis.	175
D.1	Microscope picture of a VCSEL, vertically positioned optical fibre and RF probe.	177
D.2	Microscope picture of a lumped MZM used for measurements in this thesis, with RF probe and optical fibres visible.	178

List of Tables

1.1	Mapping of the MSB and LSB to binary-coded and Gray-coded PAM4 constellations.	14
1.2	Sample EDB formation for a PRBS4 sequence.	17
3.1	Fit parameters for boron and phosphorous to be used in the Caughey-Thomas model.	77
3.2	Logic table for the four PAM levels using the switched capacitor bank.	86
4.1	Circuit values obtained from fitting of the measured S11 parameter from the lumped MZM for various bias voltages. . . .	131
4.2	3 dB frequency cut-off due to optical transit time effect for $n_g = 3.84$	135

I, Brian Murray, certify that the work I am submitting is my own and has not been submitted for another degree, either at University College Cork or elsewhere. All external references and sources are clearly acknowledged and identified within the contents. I have read and understood the regulations of University College Cork concerning plagiarism and intellectual property.

Brian Murray

Acknowledgements

The path through a doctorate is not straightforward. It presents twists and turns with the ending often obscured. Yet a destination cannot exist without the journey, and a journey is made memorable by the companions along the way.

In particular, I would like to thank my supervising team: Prof Paul Townsend, who taught me to extract the core meaning of results; Dr Giuseppe Talli, who taught me to structure research; and Dr Cleitus Antony, who taught me to always ask *why*? The Photonic Systems Group at Tyndall National Institute is a fantastic team in which to hone your craft; the surroundings, equipment and the people are world class. To everyone within and outwith the lab, I am grateful for the productive discussions and good times. Tyndall will always be close to my heart.

My time at Tyndall would not have been possible without funding from Science Foundation Ireland, to whom I extend my sincere thanks.

I would also like to thank my *viva voce* examiners for their fair comments and valuable insight: Prof Frank Peters of University College Cork and Prof Tony Kelly of University of Glasgow.

To my friends along the way, thank you for helping me keep things in perspective. To Niamh, Shane, Justin, Alison, Mohamad, Ludovic, Meysam, Mehrdad and all the lunch crew, thank you for the distractions and time well spent.

To Sean, a true Anglo-Kerryman and a truer friend, thank you for the much-needed breaks from photonics. To Steve, thank you for your staunch friendship and your valued advice, both in science and general. Thank you both for the laughs and late nights of games and pizza. I count myself lucky to have had such a consistent group of friends over the years.

To my partner, Máiréad, thank you for walking this path hand-in-hand with me. I am lucky to have found someone so naturally caring, intelligent and full of laughter. Near or far, you have never failed to cheer me up when I am down, through calm reason, bright smile or quiet empathy. Thank you for your patience despite the distance—each day together was worth the time apart. You have been my rock and motivation these past years, and I cannot wait for our adventures ahead.

Finally, I would like to express my heartfelt thanks to my family. I owe deep appreciation to my parents for the care and opportunities they gave me. They represent the best of parenting: open communication, quiet encouragement, consistency and a strong sense of morality. My childhood of books and adventure gave me curiosity and for this I will always be grateful. Mum, thank you for teaching me that if something is worth doing, it's worth doing right. This phrase rings in my head when choices are before me. Dad, thank you for teaching me to look before I leap. This remains a work in progress. Hugh, thank you for being the best of brothers and friends. No matter the years, we will always be two boys building forts in the garden.

Abstract

The boom in worldwide internet connectivity and cloud services has caused unprecedented need for high-bandwidth connections between and within data centres. Silicon photonics is becoming the platform of choice to provide low-cost, large-volume production of future optical transceivers. However, the scale of modern data centres introduces challenges of speed, reach and, crucially, energy consumption for these devices. Silicon photonic Mach-Zehnder modulators (MZMs) are one possibility for providing electrical-to-optical conversion at the transmit side of such fibre-optic links. In this thesis, comprehensive investigation is carried out into *lumped* MZMs, specifically, as their unterminated, capacitive load holds promise for lower power consumption than more typical travelling-wave MZMs with resistive terminations.

Detailed characterisations and simulations of dual-drive silicon photonic lumped MZMs are made to investigate the key trade-off of modulation bandwidth and drive voltage. Drivers with low source impedance are investigated as a means of boosting lumped MZM bandwidths, while advanced modulation formats such as four-level pulse-amplitude modulation (PAM4) and electrical duobinary modulation (EDB) are also leveraged to provide more spectrally-efficient signals. In particular, experimental demonstration is made of a novel low-impedance, switched-capacitor PAM4 driver for a lumped MZM in a 30 Gb/s silicon photonic link over 10 km of optical fibre. Simulations are carried out to optimise the bias and doping levels of lumped MZMs used with such drivers. Predistortion methods are investigated through experiments and simulations as alternative ways to increase the bandwidth. A simple first-order FIR filter is shown experimentally to enable 25 Gb/s NRZ modulation with a low-bandwidth MZM, while more optimised precompensation enables 50 Gb/s PAM4 and EDB. Finally, simulations using an accurate equivalent circuit model for the lumped MZM demonstrate the potential for a well-designed driver with lowered source impedance and controlled amounts of inductive peaking to reduce the need for transmitter-side precompensation.

Information moves, or we move to it. Moving to it has rarely been popular and is growing unfashionable; nowadays we demand that the information come to us.

NEAL STEPHENSON

Chapter 1

Photonic Systems and Data Centres

1.1 Introduction

This chapter serves as an introduction to optical communications, discussing fibre optic links, data centre trends, silicon photonics and relevant modulation formats that will be used later in the thesis. With these fundamentals in place, the later chapters of the thesis can then focus on the goal of this PhD: exploring methods of optimising lumped silicon photonic MZMs for higher bitrates and lower power consumption.

1.2 Context and Motivation

In physical cosmology, the photon epoch is the period from roughly 10 seconds to 370,000 years after the Big Bang, at which point photons dominated the energy density of the universe. Now, billions of years later, it can be argued that society has entered into a *photonics* epoch. Driven by the explosive growth of the internet in both speed and reach, photonics has become fundamental to communication, with optical fibre forming not just the backbone but the entire bone structure of global information links.

Figure 1.1 demonstrates this on a continental scale, showing that practically every major landmass is connected via multiple undersea fibre-optic cables. However, these cables form only a small fraction of the total fibre deployed worldwide. In fact, an attempt to plot the vast mesh of optical fibres that are laid on and under land would simply colour in most countries at this scale. The infrastructure that enables the internet can be divided into three hierarchical levels which

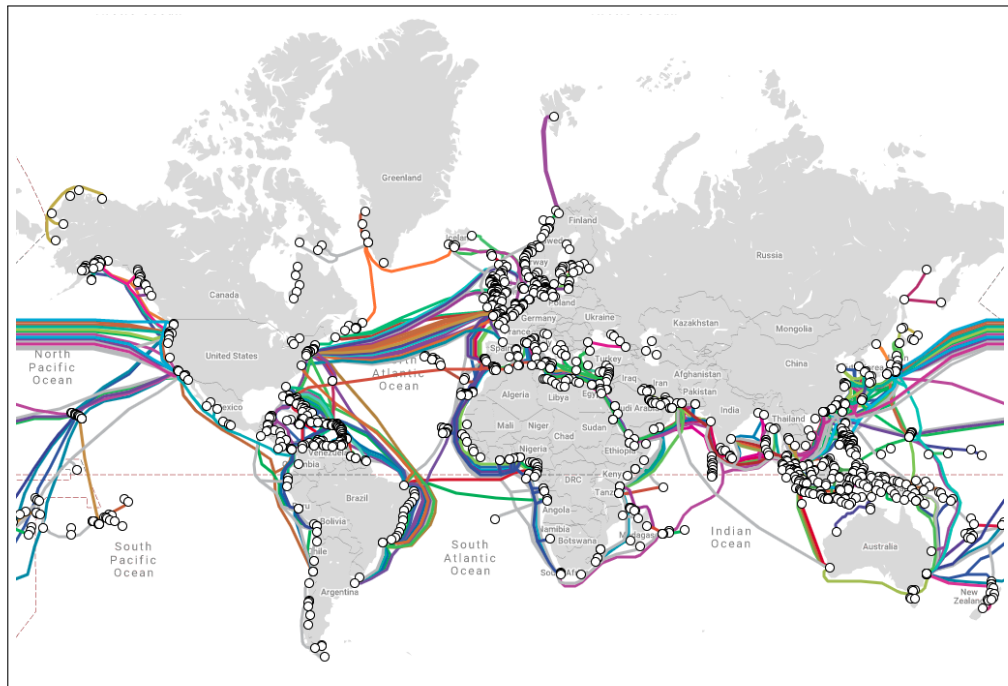


Figure 1.1: A map of global submarine optical fibres [1].

differ in terms of their geographic coverage and data traffic capacity. Ranging downwards in scale, these are the core network, the metro network, and the access network. The core network consists of links that span up to thousands of kilometres, connecting large cities within a country, between countries, or between continents. The metro network is comprised of links ranging from tens to hundreds of kilometres which link lesser nodes to the nearest core node. Finally, access networks form the capillary-like distribution system that reaches over the last tens of kilometres to connect end customers to the internet.

This infrastructure is constantly expanding. The latest Cisco annual internet report white paper [2] reveals that 4.5 billion people use the internet as of 2020, with a compound annual growth rate (CAGR) of 6 % from 2018 to 2023. This growth rate has flattened somewhat in recent years due to saturation of internet access within the population. However, the number of device connections is still increasing massively (10 % CAGR) due to users owning multiple devices and the rapid rise of machine-to-machine connections (18 % CAGR) for the internet-of-things (IoT). In addition to greater numbers of connections, the demand for faster data speed is also on the rise. Average global broadband bitrates are set to more than double over the years 2018 to 2023, from 45.9 Mbit/s to 110.4 Mbit/s—a CAGR of 20 %. This growth has caused widespread deployment of optical fibre at all network levels to cope with bandwidth demands. Indeed, fibre to the home

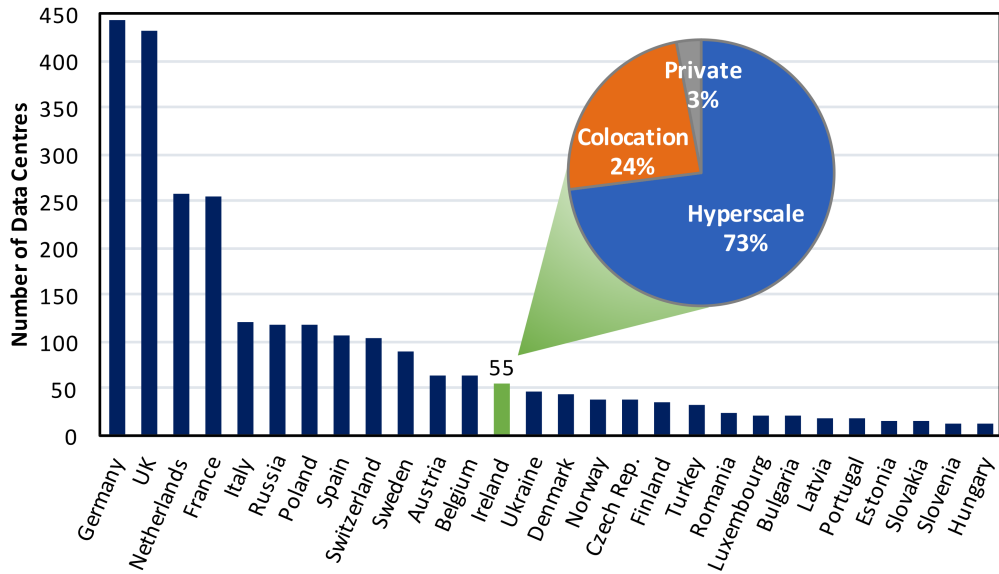


Figure 1.2: Number of data centres in countries in the European region as of 2019. Inset: breakdown of Irish data centres into private, colocation and hyperscale categories as a percentage of total power capacity. Data combined from [4, 5].

(FTTH) or to the building (FTTB) is becoming a reality for household consumers, with coverage greater than 90 % in some European countries like Spain, Portugal and Sweden (although others lag far behind) [3].

There is a missing piece in this description of how the internet is placed at our fingertips, however. To be available for distribution through the network, the information must be stored somewhere safe and readily accessible. *Enter the data centre.*

A data centre is essentially a dedicated building, space within a building, or group of buildings used to house computer servers and associated equipment that facilitate storage of and access to large quantities of data. The origins of data centres will be discussed slightly later, in Section 1.2.1, but first we will look at their setting in the current global network. Figure 1.2 shows an estimate of the number of data centres in a selection of countries in the European region as of 2019. Ireland, as a close-to-home example, hosts a moderate 55 data centres, significantly less than countries such as Germany and the UK, although as of the first quarter of 2020 there are a further 12 Irish centres under construction with 26 more seeking planning permission [6]. However, what is particularly interesting is not necessarily the total number but rather the type and scale of centres present. Data centres can be broadly classified as private (run by an

individual enterprise for its own use), colocation (lease data centre facilities and services to other enterprises) and hyperscale (owned and operated by a large company to support the services it offers, e.g., Microsoft, Google and Facebook). Inset in Figure 1.2 is the breakdown of Ireland’s current data centre types in terms of total power capacity. Noticeably, private enterprise data centres form only 3 % of this, in contrast to 97 % for the combined colocation and hyperscale categories. This is to be expected, given that there is a global trend away from in-house data centres towards more scalable and efficient leasing of larger, shared facilities [7].

It should be noted from above discussion that the measure of scale used for data centres is power (typically in the order of tens of MW) rather than floor space or server numbers. This is due to the dominant role it plays in the operational costs of such facilities—standing at approximately 35 % in 2017 [7]. The electricity costs of servers, links and switching of connections can account for up to half this power consumption, with another quarter caused by the cooling of this equipment [8]. Temperature control is a key concern for data centres due to equipment lifetimes. As server densities, quantities and speeds rise, the amount of heat they generate is increasing. As a rule of thumb, from 20 °C onwards, every 10 °C increase in a server’s temperature leads to a 50 % increase in failure rate [7].

Energy consumption from data centres is significant on a global scale, generally cited to account for approximately 1 % of the world’s energy use [9, 10, 11]. The net power consumption from data centres is often claimed to have doubled over the last decade (2010-2020) with expectations to triple or even quadruple within the next ten years [12, 13], based on the assumption that data centres’ energy consumption can be extrapolated according to predicted demand for their services. Thankfully, a more recent 2020 report [14] has shown that, although data centre energy use has increased slightly since 2010, this growth has been largely decoupled from the massive growth in data centre compute instances over the same period. This has been due to a concentrated effort to reduce energy consumption in aspects of data centre infrastructure such as cooling and power provision in order to offset increases in IT device energy use. Despite these improvements, it is clear that energy efficiency remains of paramount importance for data centre communication links—optimisation of cooling and similar factors is potentially a finite energy efficiency “resource”, and may not be sufficient on its own to facilitate future data centre computing growth. One of the oft-touted targets for the average energy needed to send information within a data centre

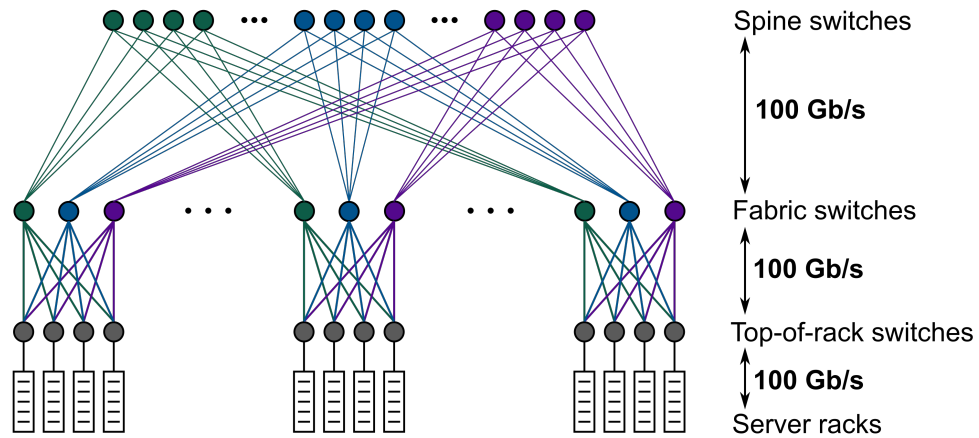


Figure 1.3: A typical data centre architecture as described by Facebook, showing 100 Gb/s link speeds. Adapted from [8, 16, 17].

is 1 pJ for every bit [15]. The reason for such a low target (and the difficulty in achieving it) can be appreciated when looking at the transmission link structure within a data centre.

Figure 1.3 shows a simplified schematic of a data centre network architecture as described by Facebook, which follows a layered design known as a leaf-and-spine architecture. This example is chosen as representative of a typical hyperscale data centre. Connections to servers in a given rack are managed via top-of-rack (TOR) switches, also known as leaf switches. These are connected in turn to higher layers in the architecture via fabric and spine switches. The design is intended to minimise the path length and switching instances needed to link any two servers in the centre. This is of key importance due to the boom of cloud computing and data analytics involving large data sets. Huge amounts of data now flows between servers (known as east-west traffic), not just between a given server and the end user (known as north-south traffic). For example, a single HTTP request by a Facebook user in 2013 could produce nearly a thousandfold increase in internal data centre traffic, even though only a small fraction of that traffic was to be returned to the user [16]. To minimise the impact of a link failing, redundancy is built in to the connection structure; as shown by the colours in the figure, multiple paths are available between each server. The total number of links between the switches in a given data centre can reach into the hundreds of thousands [8]. The distance between servers and the TOR switches is typically under 10 m, while TOR to spine switches can be 500 m or more. Further connections can also be made to spine switches in separate data centres tens of kilometres away to allow aggregation of centres within a

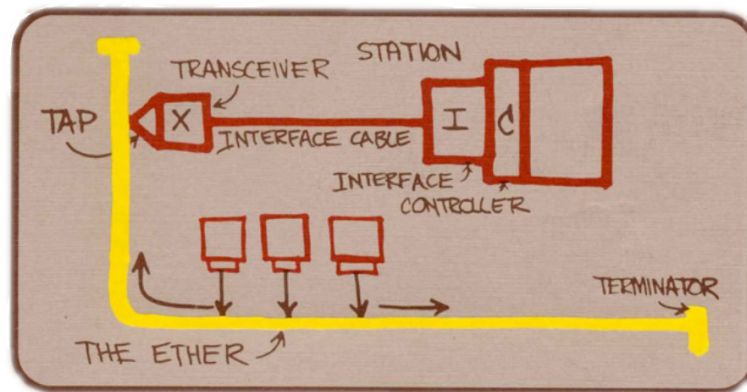


Figure 1.4: Metcalfe’s original hand-drawn diagram used to present Ethernet to the National Computer Conference in 1976 [19].

region [8]. To cope with increased demand, the speeds of connections between the servers and switches need to be upgraded. Pre-2014, connections between servers and TOR switches in Facebook’s data centres were copper based, and ran at 10 Gb/s. Communication from TOR to fabric and fabric to spine switches was at 40 Gb/s, and ran over multimode fibre (MMF) [16]. To enable the next generation of speeds, change was required in the connection medium. While links between servers and TOR switches remained over copper cable (now operating at 100 Gb/s), the multimode fibre in the higher levels was replaced with single-mode fibre (SMF) to allow connections between TOR, fabric and spine switches to also run at 100 Gb/s over distance [17].

To understand the decision to shift the optical infrastructure from multimode fibre to single-mode fibre, it worth looking at how data centre networks originally developed. This follows in the next section.

1.2.1 Data Centres: from Copper to Single-Mode Fibre

The sharing of digital information through a high-speed local area network (LAN) has its origins in the 1970s, when computing was dominated by very large and costly mainframe computers [18]. In 1973, a Xerox employee, Bob Metcalfe, wrote a memo describing his invention of “Ethernet”, a networking system that allowed computer workstations within the same building to send information to each other and to connected printers. His system was based upon a previous, slower but longer-range radio network known as ALOHAnet, which was developed at the University of Hawaii in the late 1960s to connect users on the islands with a central time-sharing computer.

In 1976, Ethernet was presented to the USA's National Computer Conference using the diagram shown in Figure 1.4, which illustrates workstations connected via transceivers that tap into a distributed linking medium. The word used to describe this medium, *ether*, was a reference back to the so-called "luminiferous aether" of 19th century science, through which electromagnetic waves had been incorrectly thought to propagate [20]. Although the connecting medium in this implementation was co-axial cable, Ethernet was viewed as medium-independent from the beginning [21]. Indeed, in the years since then, LAN connections have also been implemented on twisted-pair cabling, wireless connections, and optical fibre. The choice of medium used for a given network depends on the application space, and, among other factors, its required bitrate and link distance.

In contrast to comparatively short-reach LANs, the textbook definition of a wide area network (WAN) is a computer network spanning regions, countries or even the world [22]. The internet, our truly global WAN, is now taken for granted. However, it took the laying of undersea optical fibres such as TAT-8 in 1988 costing hundreds of millions of dollars [23] to facilitate computing between continents. The TAT-8 cable was based on single-mode fibre (SMF), with lasers and photodetectors operating at an optical wavelength of $1.3\ \mu\text{m}$, while later cables such as FLAG in the 1990s operated at $1.5\ \mu\text{m}$. These wavelengths were chosen as being optimal for long-distance transmission. In the case of TAT-8, $1.3\ \mu\text{m}$ was used due to its zero in chromatic dispersion, while the emergence of erbium-doped fibre amplifiers (EDFAs) at $1.5\ \mu\text{m}$ dictated the wavelength for later cables [24].

The choice of wavelength to be used for the optical parts of a LAN or WAN has always depended on its implementation cost and its tolerance to optical attenuation and dispersion. With this in mind, it is worth looking at how the attenuation of fibre varies across the wavelengths typically employed for telecommunications, and how the communication windows in this spectrum are chosen as a result. Figure 1.5 shows the attenuation spectrum for SMF in the early 1980s, for modern SMF, and also for a modern multimode fibre (MMF). The larger absorption towards lower wavelengths is due to Rayleigh scattering from localised microscopic variations in the material density of the fibre, a loss mechanism that causes intensity to scale with the inverse fourth power of wavelength. At wavelengths roughly above $1700\ \text{nm}$, infrared absorption is the dominant loss mechanism, causing the attenuation to rapidly rise with wavelength. The region of low-loss between the grip of these two intrinsic absorption mechanisms is punctuated by noticeable peaks. Extrinsic absorption

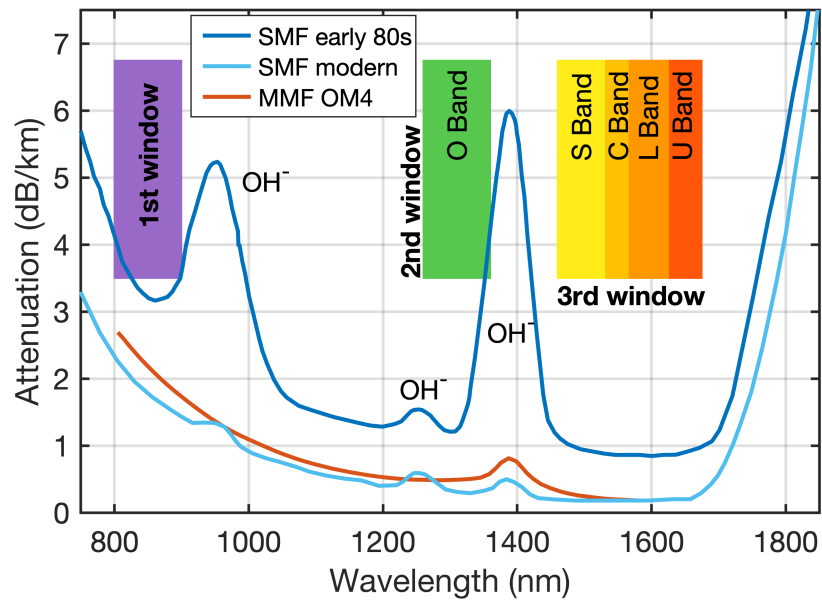


Figure 1.5: Attenuation profile and transmission windows for optical fibre. SMF data and MMF data were extracted from [25, 26], respectively, using [27].

is caused by unwanted impurities introduced during the fibre manufacturing process. These so-called water peaks are due to the presence of the hydroxide ion, OH^- , whose vibrational states cause spikes in absorption at 950 nm, 1240 nm and 1580 nm. Figure 1.5 shows clearly that fibre attenuation has reduced over time due to improvements in manufacturing, and also demonstrates how the removal of residual water vapour in the process has almost entirely suppressed the water peaks.

As can be seen from Figure 1.5, the attenuation profile of modern OM4-quality MMF is similar to modern SMF. In fact, MMF has largely matched SMF over the years in terms of losses, but is more restricted in reach due to its inherent modal dispersion. Nevertheless, the optical links used in LANs have until recently been dominated by MMF. This is due to its large core diameter, which simplifies coupling to low-cost optical components such as light-emitting diodes (LEDs) and vertical-cavity surface-emitting lasers (VCSELs). Historically, such devices were readily available at 850 nm, which falls in a region of relatively low-loss known as the first transmission window or the 850 nm band. For longer-distance connections, as mentioned above, the main choices of wavelength have been the second transmission window from 1260 nm to 1360 nm, called the “original” or O band, and the “conventional” or C band from 1530 nm to 1565 nm, located in the third transmission window. Over the years, other telecom bands have also been introduced within this third window, forming the S band (short wavelengths), L

band (long wavelengths) and U band (ultra-long wavelengths). Although it is not shown in Figure 1.5, the suppression of the primary water peak has allowed the introduction of the E band (extended) that bridges the second and third telecom windows. However, this is less used than other bands, as legacy fibre can still exhibit a considerable water peak.

The rise of data centres in the modern world has placed increasing demands on network capabilities. As a result, LAN speeds, reach and the choice of transmission medium have had to adapt accordingly. Ethernet's prototype iteration worked at a bitrate of 2.94 Mb/s over a maximum distance of 1000 m [20], while a faster 10 Mb/s version over 500 m links followed in 1980. Since 1985, wired Ethernet has been standardised by the IEEE 802.3 committee. The progress of these standards for various types of copper cabling, MMF and SMF is illustrated in Figure 1.6, a viewing of which conveniently reveals trends in data centre technology. The first graph, Figure 1.6(a), shows the target bitrate of Ethernet standards over the years. As could be expected, the rates trend upwards. MMF was first introduced to the standard in 1990 at 10 Mb/s, before SMF followed in 1998 for the 1 Gb/s standard. Notably, barring the introduction of some lower-bitrate copper standards for automation and automotive applications in the late 2010s, optical fibre and copper have matched each other stride for stride towards 100 Gb/s and 400 Gb/s standards. However, looking at link distance in the second graph, Figure 1.6(b), it is clear that the increase in bitrates has led to copper and MMF being used over shorter and shorter links. At such high frequencies, copper cables suffer greatly from signal attenuation, while MMF is intrinsically restricted by modal dispersion. The message is clear: as data centres grow larger and become increasingly connected to one another, SMF has become the working solution for their intraconnects and interconnects. Figure 1.6(c) emphasises this further by showing the bitrate-distance product of the standards. The trend lines slope upwards for all three media, but SMF has remained consistently higher in bitrate-distance product than the other options. In short, the foreseeable future for data centre links is rooted in single-mode fibre.

1.2.2 Silicon Photonics

Around the same time that data centres have been switching from MMF to SMF, a change has also arisen in the material platform for the transmitters and receivers used at either end of the links. There are four main material technologies

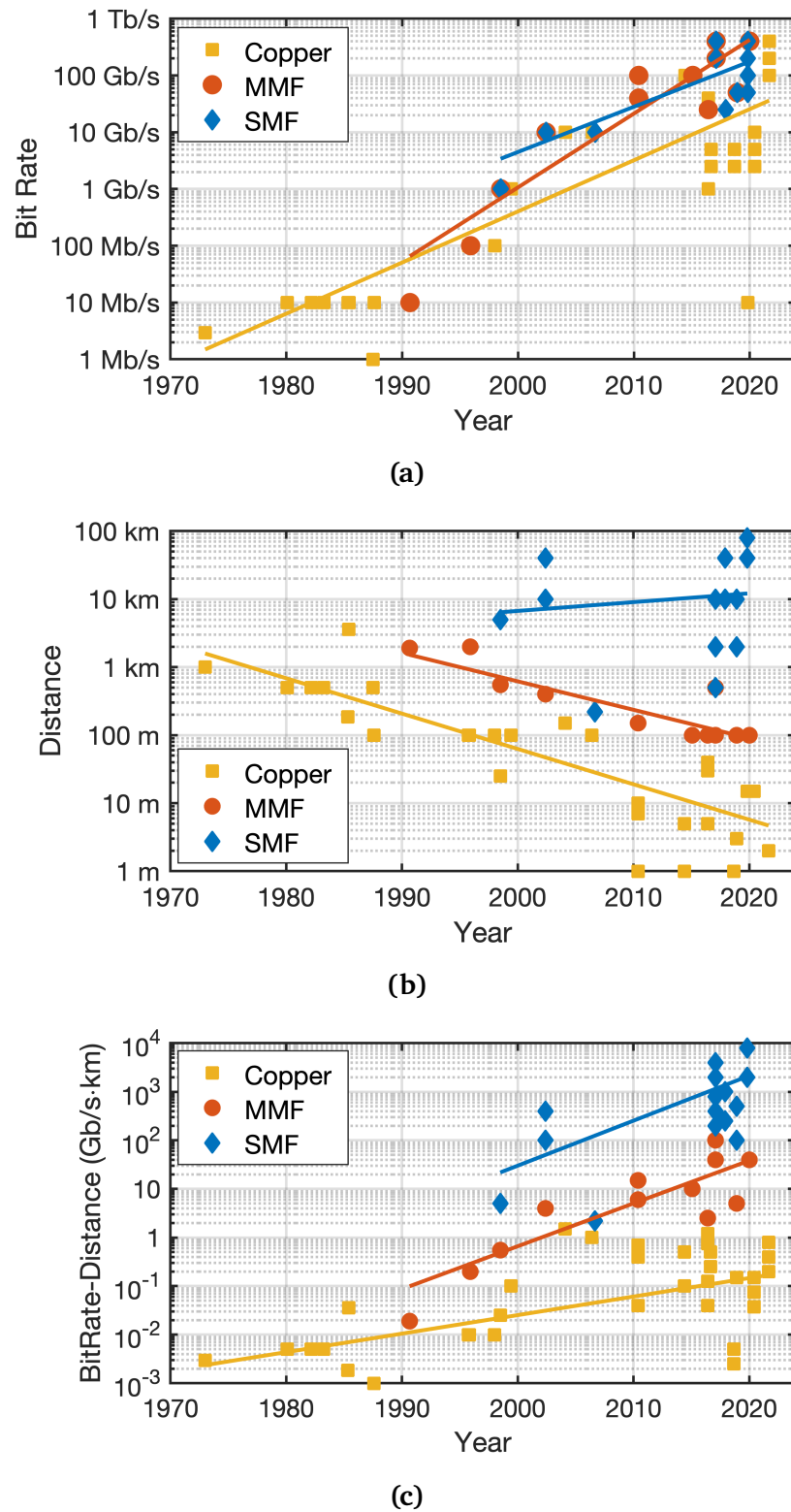


Figure 1.6: Progression of Ethernet standards in terms of (a) bit rate, (b) distance, and (c) bit-rate-distance product, compiled from [20, 28, 29, 30].

used to fabricate optical components for transmitters and/or receivers: lithium niobate (LiNbO_3), indium phosphide (InP), gallium arsenide (GaAs) and silicon (Si). Lithium niobate and indium phosphide have largely secured the long-haul market for inter-country and intercontinental transmissions over the core network. Gallium arsenide is used (among other applications) for VCSELs, which have been widely deployed until recently for short-distance links in data centres with MMF, as mentioned in Section 1.2.1. Silicon photonics (SiP) is the newest of these optical technologies and is uniquely positioned to capture the data centre market for a number of reasons.

One of the primary reasons for creating photonics on a silicon platform is that it allows device manufacturers to piggyback on the enormous industry of semiconductor electronics. Complimentary metal-oxide semiconductor (CMOS) has been the foundation of the integrated circuit industry for decades, using vast quantities of silicon in the mass manufacture of chips. Due to the maturity of this material platform, silicon photonic devices can be fabricated on larger wafers with higher yield than traditional photonic integrated circuits made using InP and GaAs. The result—more devices being made simultaneously on the same wafer—allows silicon photonics fabrication facilities to operate in a more cost-effective fashion than other material processes.

Silicon photonics has also benefited from a move to wavelength-division multiplexing (WDM) schemes in data centres, which has been necessary to reach standards of 100 Gb/s and beyond. In WDM, data is split onto multiple carrier wavelengths with each stream typically operating at 25 Gb/s or 50 Gb/s, before being transmitted down the same fibre and recombined at the receiver. Although WDM could be adopted for VCSELs, achieving tight wavelength selectivity with these devices is not simple [31]. Furthermore, VCSELs are restricted in their modulation bandwidths due their operation as directly-modulated lasers [32]. These limitations, in combination with the shift to SMF ending the benefits of looser multimode alignment tolerances, mean that VCSELs have fallen out of favour for WDM solutions in data centres. Silicon photonics has the ability to co-integrate various photonics building blocks reliably on a single chip; such photonic integrated circuits (PICs) are highly beneficial for WDM systems which require multiple lasers and modulators for each wavelength, along with a multiplexer, demultiplexer and grating couplers.

Silicon photonics is eminently suitable for PICs, but its potential for integration is not just restricted to the optical domain. Richard Soref, often called the

founding father of silicon photonics due to his seminal papers on the field [33, 34], outlined well its suitability for optoelectronic integrated circuits, which involve the fabrication and integration of optics and electronics on the same substrate [35]. There are numerous advantages to such a single-chip amalgamation of photonics and electronics, most notably the potential reduction of input-output (I/O) bottlenecks that are typically formed by the poor scaling of electrical I/O with hikes in transmission speeds.

However, there is a key disadvantage of silicon as a material platform for photonics: due to its indirect bandgap, pure silicon does not lase. Without the ability to monolithically grow a laser on the silicon chip, any function that requires a light source will necessitate a hybrid approach, i.e., a laser created on a different material platform must be externally coupled or else bonded onto the silicon using heterogeneous integration. Naturally, this adds some level of fabrication complexity and cost to silicon transceivers, although several CMOS-compatible processes have been demonstrated as potential solutions [36, 37]. The ultimate fix for silicon's laser problem is an ongoing topic of photonics research that is beyond the scope of this thesis, thus any lasers used as optical sources in this work are all commercial, bench-top devices.

As will be discussed in more detail in Section 2.3, silicon's lack of a strong linear electro-optic effect means that it requires doping with free carriers to facilitate modulation and, hence, data transmission. The challenge in implementing this method of modulation is balancing high-speed performance against electrical efficiency. Modulators must be of sufficient bandwidth to prevent intersymbol interference (ISI) introducing bit errors into the signal. In general, higher-bandwidth modulators need to be shorter in length and so require more voltage to achieve the same optical signal swings as lower-speed devices. This is an issue in practical data centre applications, where drive voltages must be kept low to reduce power consumption and prevent overheating and damage to equipment¹. The efficient use of bandwidth and voltage in silicon modulators is an important part of the work in this thesis.

¹An enjoyable anecdote that cautions simply using *more voltage* to send a signal is that of Dr. Wildman Whitehouse and the first transatlantic telegraph cable. In 1858, frustrated at the poor signal-to-noise ratio of messages between the continents, Whitehouse used his patented induction coils to force 2000 V down the line. In doing so, he blew a hole in the pioneering cable's insulation somewhere between Ireland and Newfoundland, rendering the entire system permanently out of order within its first four weeks of operation [38].

1.3 Multilevel Modulation Formats

Non-return-to-zero on-off keying has been the de facto choice of modulation formats in most optical communication systems for decades due to its simplicity of implementation. The term *on-off keying* (OOK) refers to the fact that the signal has binary information coded directly in pulses of maximum and minimum amplitude. The signal is of *non-return-to-zero* (NRZ) format in that it is not dropped to zero amplitude in the case of consecutive high-amplitude symbols. In this thesis, as is commonly done in optical communications literature, the term NRZ will be used to refer specifically to NRZ-OOK. A typical NRZ signal bit stream and eye diagram are shown in Figure 1.7(a). As it consists of only two amplitude levels, the bit content in an NRZ signal can be easily decoded at the receiver by a single comparator or slicer. However, the disadvantage of NRZ is its relatively low spectral efficiency; it carries only a single bit of information in every pulse or symbol. With transceiver bandwidth typically at a premium in data-centre settings, the use of more spectrally efficient modulation formats is of considerable interest.

1.3.1 Four-Level Pulse Amplitude Modulation

In general, the term PAM refers to a family of intensity modulation formats which all follow the same principle: individual bits of information can be combined with other bits to form symbols, and then transmitted in the form of a signal with multiple amplitude levels, each of which corresponds to a given symbol. If the signal has M amplitude levels, each symbol is then comprised of $\log_2(M)$ bits. Naturally, if $M = 2$, the resulting PAM2 signal is no different to an NRZ signal. By using more than two amplitude levels, the required symbol rate needed to transmit a signal at a given bitrate is reduced, i.e., it decreases by a factor of $\log_2(M)$. This is beneficial in that it reduces the clock-rate and bandwidth requirements of the transmission link.

For example, to form a PAM4 signal, through which the required symbol rate is halved, two binary streams are combined bit-by-bit. In this pairing process, the corresponding bits from each stream are called the most-significant bit (MSB) and least-significant bit (LSB), respectively. Constellation maps such as the binary coded and Gray coded examples shown in Table 1.1 are used to determine bit content of the four symbols or amplitude levels in the signal². Gray coding differs

²The convention in this thesis will be to refer to the amplitude levels of the PAM4 signal as

Table 1.1: Mapping of the MSB and LSB to binary-coded and Gray-coded PAM4 constellations.

PAM Level	Binary MSB/LSB	Gray MSB/LSB
Four	11	10
Three	10	11
Two	01	01
One	00	00

from binary coding in that the pairs encoded in neighbouring symbols differ only by one bit, and so is typically favoured to reduce the impact of symbol errors on the signal quality (comparing levels two and three in the table shows how mixing up these symbols at the receiver would lead to a double bit error in the binary code, as opposed to a single bit error in the Gray code).

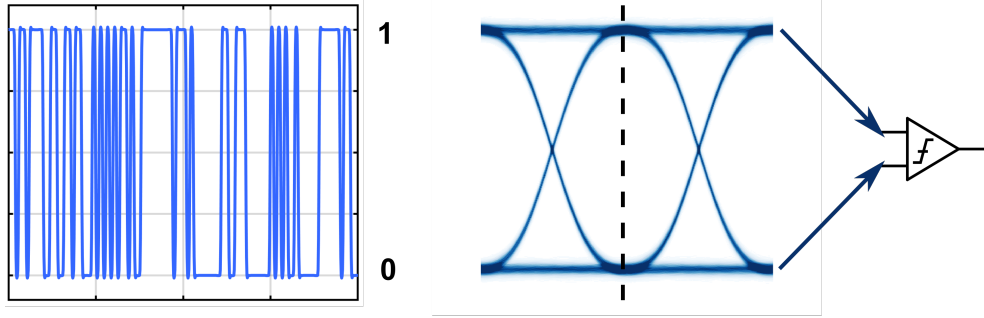
In theory, there is no limit to the number of amplitude levels that can be introduced to a PAM signal, provided a suitably sensitive receiver can distinguish one level from another at the end of the link. However, in practice, the linearity requirements and reduced noise tolerance introduced by the use of multiple levels means that the work in this thesis focusses on PAM4 and, to a lesser extent, PAM8. A typical PAM4 waveform and eye are shown in Figure 1.7(b).

1.3.2 Electrical Duobinary Modulation

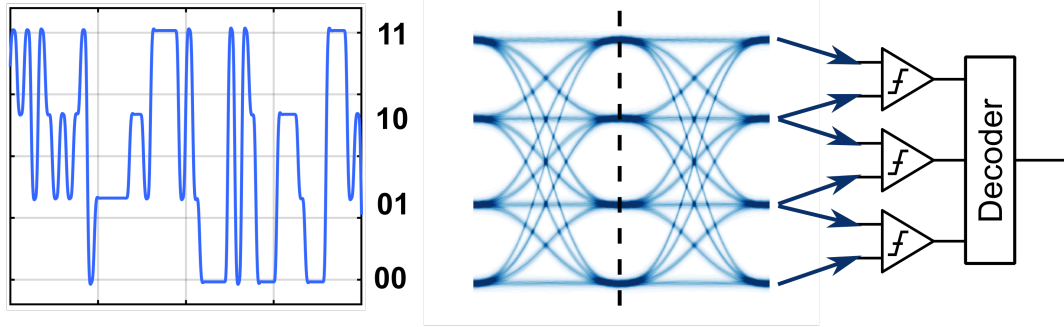
An alternative method of achieving multilevel modulation is to employ partial-response signalling, where controlled ISI is introduced to the transmission system. The resulting spreading of each pulse into its neighbours means that the pulse durations are longer, and, hence, the frequency spectrum of the signal is narrower. Crucially, as this ISI is added in a controlled fashion, it can be removed from the signal at the receiver. The reduction of the spectrum through partial-response signalling places reduced bandwidth requirements on the components of the system. Although partial-response signalling can be employed with a PAM signal of any number of levels, this thesis will focus only on the case where it is used for PAM2 or NRZ, for which the resulting signal is known as an electrical duobinary (EDB) signal. While EDB does not reduce the symbol rate needed for a given bitrate (contrary to PAM4, for example), the reduced frequency spectrum of the

levels one, two, three and four. This is a convenient labelling, but does lead to the slightly odd situation where binary 00 refers to "one" rather than "zero", etc.

(a) On-off keying, non return to zero (NRZ)



(b) Four-level pulse-amplitude modulation (PAM4)



(c) Electrical duobinary modulation (EDB)

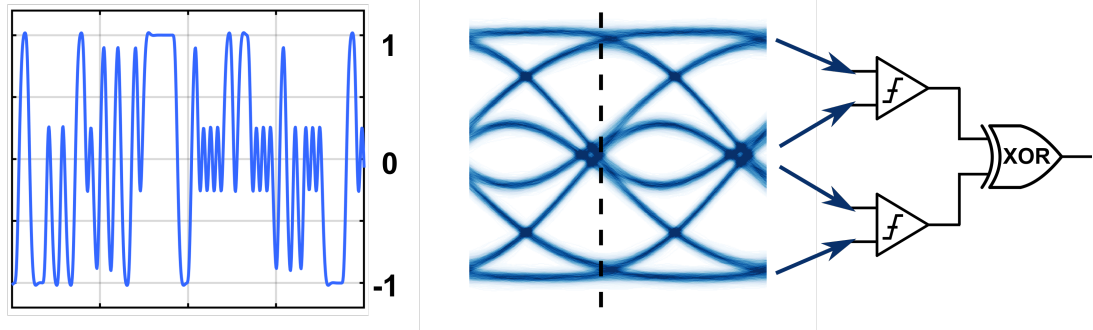


Figure 1.7: Simulated waveforms and corresponding eye diagrams for (a) NRZ, (b) PAM4 and (c) EDB. The ideal sampling point for each eye diagram is shown as a dotted line. Also shown are block diagrams of the receiver logic.

duobinary signal makes it more resilient to bandwidth impairments than an NRZ signal of the same bitrate.

The duobinary encoding which introduces the controlled ISI can be accomplished through a delay-and-add filter which sums each bit with the previous bit, or through a low-pass electrical filter with a 3 dB bandwidth of approximately one quarter the symbol rate. The k^{th} symbol c_k in the resulting duobinary signal after this filtering is then

$$c_k = a_k + a_{k-1}, \quad (1.1)$$

where a_k and a_{k-1} are the corresponding bits in the input data. In this case, to decode the original information, the receiver simply reverses the duobinary encoding process through a feedback process:

$$\hat{a}_k = \hat{c}_k - \hat{a}_{k-1}, \quad (1.2)$$

where we have placed hats over the symbols $\hat{a}_k$, $\hat{a}_{k-1}$ and $\hat{c}_k$ to indicate that these correspond to bits as interpreted at the other end of the link, where there is always the risk of an error being present. This distinction is particularly important as the presence of feedback in the decoder introduces the possibility of unbounded error propagation; an error in one bit will transfer to the process of decoding the next bit through Equation 1.2, and so on. To avoid this, an additional precoding step is generally performed at the transmitter as shown in Figure 1.8, where each precoded bit b_k is formed from the input signal through

$$b_k = (a_k - b_{k-1}) \mod 2, \quad (1.3)$$

which is equivalent to an XOR logic operation. The duobinary signal after the usual subsequent filtering is now

$$c_k = b_k + b_{k-1}. \quad (1.4)$$

After some simple algebra and substitution of Equations 1.3 and 1.4, it can be shown that receiver needs only take the modulo-2 of the sampled symbols (with an XOR logic gate) to decode the duobinary signal back to binary through

$$\hat{a}_k = \hat{c}_k \mod 2, \quad (1.5)$$

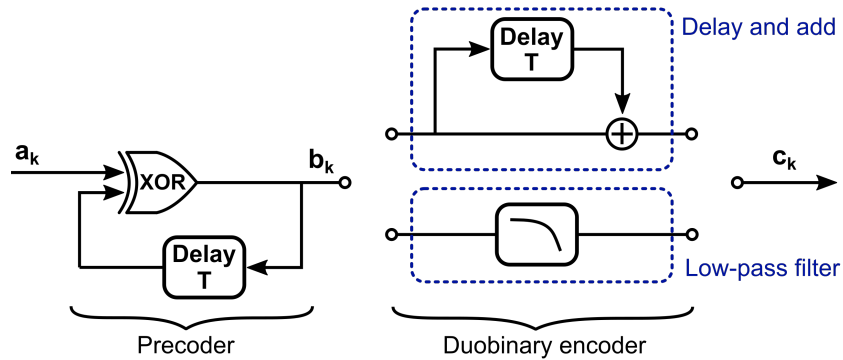


Figure 1.8: Block diagram of a duobinary transmitter comprising of the precoder and the two methods of duobinary encoder.

Table 1.2: Sample EDB formation for a PRBS4 sequence.

$a_k :$	1	1	1	1	0	1	0	1	1	0	0	1	0	0	0
$b_k :$	1	0	1	0	0	1	1	0	1	1	1	0	0	0	0
$c_k :$	1	1	1	1	0	1	2	1	1	2	2	1	0	0	0
$\hat{a}_k :$	1	1	1	1	0	1	0	1	1	0	0	1	0	0	0

where we have added a hat over $\hat{c}_k$ to emphasise that this is c_k as received at the end of a link which may have introduced a bit error. Crucially, however, the precoding in Equation 1.4 means that the feedback step has already taken place at the transmitter, so a bit error at the receiver will not lead to error propagation.

An example of the steps involved in EDB encoding (with prior precoding) followed by the decoding at the receiver is shown in Table 1.2 for a PRBS4 sequence. The symbols in the EDB sequence as calculated directly are $c_k \in \{0, 1, 2\}$., but are often equivalently labelled as $c_k \in \{-1, 0, 1\}$. Comparing c_k and $\hat{a}_k$, it is clear that the outer duobinary symbols correspond to logic 0 of the input data and the inner symbol corresponds to logic 1. Notably, not all possible symbol transitions are present for the sequence of c_k : duobinary signals always move in step, never jumping from from one outer rail to another. A typical EDB waveform and eye obtained from the low-pass filter encoding method are shown in Figure 1.7(c).

To be clear, in this thesis, the performance of NRZ, PAM4 and EDB signals will typically be compared at the same *bitrate*, with symbol rates chosen appropriately for each format. For example, a comparison of the three formats at 50 Gb/s would use 50 Gbaud NRZ, 25 Gbaud PAM4 and 50 Gbaud EDB. To meet the Nyquist rate, the transmitter and receiver circuitry would then operate at a minimum sample rate of 100 GS/s for NRZ, 50 GS/s for PAM4 and 100 GS/s for EDB.

1.3.3 Bit Error Rate

The ultimate metric for the quality of a signal under given conditions is the bit error rate (BER), defined as the ratio of the number of incorrectly received bits to the total number of received bits in a certain time frame. In this section, a standard derivation [39] of an analytic expression for the BER of a binary amplitude-modulated signal is given, which is then extended to multilevel amplitude-modulated signals.

In a binary system that has modulated symbols 0 and 1 that are equally likely to

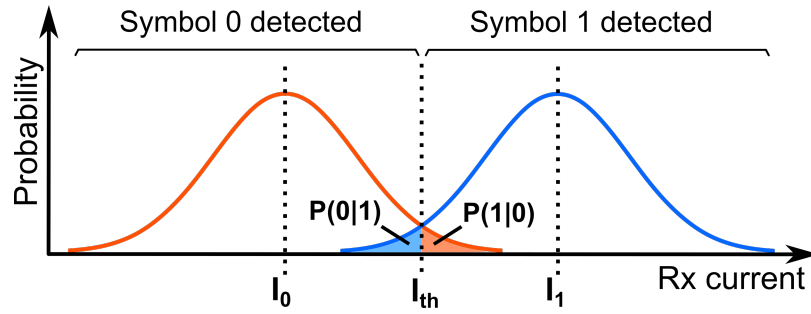


Figure 1.9: Normal distribution of binary symbol samples and the receiver decision regions.

occur, the BER can be expressed as

$$BER = \frac{1}{2} [P(0|1) + P(1|0)], \quad (1.6)$$

where $P(0|1)$ and $P(1|0)$ are, respectively, the probability of a transmitted 1 being incorrectly detected as a 0, and a transmitted 0 being incorrectly detected as a 1. The functional form of these conditional probabilities depends on the statistics of the noise sources responsible for fluctuations in the signal photocurrent I generated by the incident light at the receiver. Under the assumption that the system is limited by thermal noise, these fluctuations are well described by Gaussian statistics with mean μ and variance σ^2 , and so follow a normal distribution:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x - \mu^2}{2\sigma^2}\right). \quad (1.7)$$

To distinguish received symbols from each other, the receiver must set a decision threshold I_{th} for the current, below which any sampled value is taken as a 0 and above which any sampled value is taken to be a 1, as shown in Figure 1.9. If I_0 and I_1 are the mean of the respective photocurrents at the receiver for symbols 0 and 1, respectively, and σ_0^2 and σ_1^2 are their respective variances, the conditional probabilities can be expressed as

$$P(0|1) = \frac{1}{\sigma_1\sqrt{2\pi}} \int_{-\infty}^{I_{th}} \exp\left(-\frac{(I - I_1)^2}{2\sigma_1^2}\right) dI = \frac{1}{2} \operatorname{erfc}\left(\frac{I_1 - I_{th}}{\sqrt{2}\sigma_1}\right), \quad (1.8)$$

$$P(1|0) = \frac{1}{\sigma_0\sqrt{2\pi}} \int_{I_{th}}^{\infty} \exp\left(-\frac{(I - I_0)^2}{2\sigma_0^2}\right) dI = \frac{1}{2} \operatorname{erfc}\left(\frac{I_0 - I_{th}}{\sqrt{2}\sigma_0}\right). \quad (1.9)$$

Here, for convenience, we have introduced the complimentary error function, defined as

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} \exp(-y^2) dy. \quad (1.10)$$

Then, by substituting Equation 1.8 and Equation 1.9 into Equation 1.6, the BER can be expressed as

$$BER = \frac{1}{4} \left[\operatorname{erfc} \left(\frac{I_1 - I_{th}}{\sigma_1 \sqrt{2}} \right) + \operatorname{erfc} \left(\frac{I_{th} - I_0}{\sigma_0 \sqrt{2}} \right) \right]. \quad (1.11)$$

Clearly, Equation 1.11 depends on the decision threshold I_{th} , which should be chosen to minimise the BER. This happens when

$$\frac{(I_{th} - I_0)^2}{2\sigma_0^2} = \frac{(I_1 - I_{th})^2}{2\sigma_1^2} + \ln \left(\frac{\sigma_1}{\sigma_0} \right). \quad (1.12)$$

In most cases of practical interest, σ_0 and σ_1 are similar enough that the last term in this equation is negligible. The optimal value of I_{th} for given symbol means and variances is then

$$I_{th} = \frac{\sigma_0 I_1 + \sigma_1 I_0}{\sigma_0 + \sigma_1}. \quad (1.13)$$

Under the thermally-limited conditions at the receiver, the noise variance is independent of the average photocurrent, and so $\sigma_0 = \sigma_1$. In this case, the optimal decision threshold is $I_{th} = (I_1 + I_0)/2$, which corresponds to setting the decision threshold in the middle.

Regardless of whether the signal is thermal-noise limited or not, the quality factor Q can be introduced as

$$Q = \frac{I_1 - I_0}{\sigma_0 + \sigma_1}. \quad (1.14)$$

Then, combining Equations 1.11, 1.13 and 1.14, the BER of the binary signal is expressed simply as

$$BER = \frac{1}{2} \operatorname{erfc} \left(\frac{Q}{\sqrt{2}} \right). \quad (1.15)$$

This analysis can also be extended to multilevel pulse-amplitude modulation (PAM) formats with M symbol levels [40]. Similarly to Equations 1.8 and 1.9, the probability of receiving symbol j when symbol i is transmitted is

$$P_{ij} = \frac{1}{2} \operatorname{erfc} \left(\frac{I_{th,j} - I_i}{\sigma_i \sqrt{2}} \right) + \frac{1}{2} \operatorname{erfc} \left(\frac{I_i - I_{th,j+1}}{\sigma_i \sqrt{2}} \right), \quad (1.16)$$

where I_i is the receiver photocurrent at symbol i , $I_{th,j}$ is the threshold current for symbol j , and σ_i^2 is the variance of the noise current for symbol i . As each symbol now can represent multiple bits, the more general symbol error rate (SER) is

used rather than BER, and is calculated from

$$SER = \frac{1}{M} \sum_{i=0}^{M-1} \sum_{j=0, j \neq 1}^{M-1} P_{ij}. \quad (1.17)$$

Assuming that all symbol levels are equally spaced and that the various decision thresholds are equidistant from adjacent symbols, the SER can then be expressed as

$$SER = \frac{M-1}{M} \operatorname{erfc} \left(\frac{I_{avg}}{(M-1)\sqrt{2}\sigma} \right), \quad (1.18)$$

where I_{avg} is the average receiver photocurrent and we have taken $\sigma_i = \sigma$ for all symbols. For sufficiently high signal-to-noise ratios (i.e., when I_{avg}^2/σ^2 is large), it can be assumed that only errors between adjacent symbols occur. In that case, the BER for a given PAM order can be approximated as

$$BER \simeq \frac{d_{avg}}{\log_2(M)} SER \quad (1.19)$$

where d_{avg} is the average Hamming distance between the labels of symbols i and j . For example, in the case of Gray-coded PAM4, the BER can be taken as being equal to one half of the SER. Throughout this thesis, Gray coding will be used for multilevel PAM formats. In this scheme, the bit groups of neighbouring symbols differ optimally by only one bit, and hence $d_{avg} = 1$. This expression for the BER of a multilevel PAM signal will be used in the analysis in Section 3.4.2.

The calculation of BER for an EDB signal is similar to that of a PAM signal, and an excellent treatment is given in [41]. Under the same assumptions made above (where the symbol levels are equally distributed and have equal noise variance, and where errors occur only between adjacent symbols), the BER for an EDB signal reduces to

$$BER = \frac{3}{4} \operatorname{erfc} \left(\frac{I_{avg}}{2\sqrt{2}\sigma} \right), \quad (1.20)$$

This is similar to the SER as calculated in Equation 1.18 where $M = 3$, although the distribution of the bits in the outer levels of the duobinary signal affects the constant term at the start of the expression. A derivation of this expression is provided in Appendix A.

To compare the BER curves for NRZ, PAM4 and EDB, it is necessary to define suitable expressions for the average photocurrent I_{avg} and noise variance σ^2 at the receiver. In photodiodes, the generated photocurrent is directly proportional

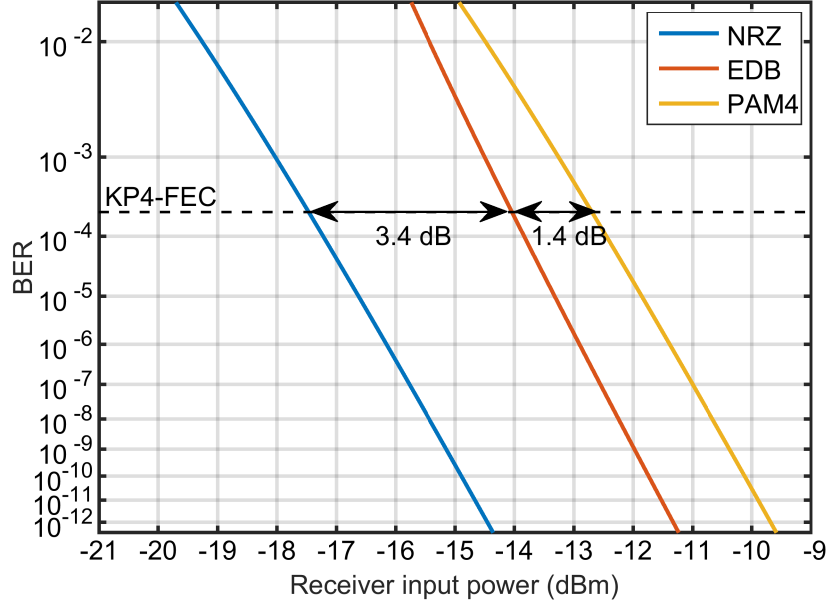


Figure 1.10: Comparison of simulated BER curves for NRZ, EDB and PAM4.

to the incident optical power. Hence, the average photocurrent is given by

$$I_{avg} = R \cdot P_{avg}, \quad (1.21)$$

where R is the responsivity of the photodiode, and P_{avg} is the average incident optical power. The variance of thermal noise at the receiver depends in particular on the electrical bandwidth of the receiver, in addition to the noise-inducing effects of any electrical amplifiers present. A suitable expression for the noise variance is then

$$\sigma_{th}^2 = \frac{4k_B T}{R_L} F_n B_{eff}, \quad (1.22)$$

where k_B is the Boltzmann constant, T is the temperature of the photodiode in Kelvin, R_L the value of the load resistor, B_{eff} is the effective electrical bandwidth of the receiver, and F_n is the amplifier noise figure [39]. However, instead of giving a value for F_n , most photoreceiver data sheets typically provide the noise equivalent power (NEP), with units of $\text{pW}/\sqrt{\text{Hz}}$. This is defined as the minimum optical input power per unit electrical bandwidth required to produce a signal-to-noise ratio of 1 [39]. The noise variance is then simply calculated as

$$\sigma^2 = R \cdot NEP \cdot \sqrt{B_{eff}}. \quad (1.23)$$

Using the above equations, the simulated BER curves for NRZ, EDB and PAM4 signals are compared in Figure 1.10, for a photoreceiver from Finisar (model

XPRV2021a) that is specified to have an NEP of $25 \text{ pW}/\sqrt{\text{Hz}}$, a responsivity of 0.625 A/W , and an effective (3 dB) bandwidth of 40 GHz. Typically, different BER curves are compared in terms of a sensitivity penalty, that is, the difference in optical power required at the receiver to obtain a certain reference BER. A signal is typically deemed to be “error-free” if the BER is less than 10^{-12} . However, most optical communication links operate using forward-error correction (FEC), which can almost entirely remove errors in a signal provided the uncorrected BER remains below a given number known as the FEC threshold, which depends on the type of FEC implemented. In this case, the reference BER level is taken as the KP4-FEC threshold, a typical FEC value for data centre settings. When compared with NRZ, the EDB and PAM4 signals exhibit sensitivity penalties of 3.4 dB and 4.8 dB, respectively. These penalties correspond to EDB eye opening being reduced by approximately a linear factor of 2 (equivalent to 3 dB) and the PAM4 eye opening reducing by a linear factor of 3 (equivalent to 4.8 dB).

This analysis of the BER holds only in the case of a system limited by thermal noise, and does not account for the effects of bandwidth limitations or fibre dispersion, which play significant roles in the high bit rates demanded by data centres. Hence, such effects will be evaluated through more involved experimental measurements and simulations throughout this thesis.

1.4 Summary

This chapter evaluates the recent trends in data centres interconnects, highlighting the recent shift from VCSELs and MMF links towards silicon photonics and SMF. An overview is also given of relevant multilevel direct-detection modulation formats. Section 1.2 places the evolution of modern data centres within the historical context of Ethernet and LAN development. The advantages (and challenges) of the silicon photonics platform for data centre links are discussed, highlighting its possibilities for low-cost, high-volume production and optoelectronic integration. Section 1.3 gives key description of PAM4 and EDB signal formats. A discussion is also given of their BER and eye diagrams, both of which will be used extensively throughout this thesis in the presentation of experimental and simulated results.

Chapter 2

Fundamentals of Silicon Photonic Mach-Zehnder Modulators

2.1 Introduction

This chapter introduces the fundamentals of silicon photonic Mach-Zehnder modulators. Such devices have been widely researched, with numerous books and theses covering the introductory topics. Therefore, parts of this chapter follow a similar structure to the excellent treatment given in [42, 43, 44]. The crucial building blocks of Mach-Zehnder modulators are discussed, from silicon waveguides and grating couplers to the various types of phase shifter technologies and drive formats, making comparison, where relevant, to the more traditional phase shifter technology of LiNbO_3 . Clear distinction is made between the design of lumped MZMs and travelling-wave MZMs, and attention is given to estimating the power consumption and drive-voltage requirements of both. The state of the art in silicon MZMs is also reviewed.

2.2 The Mach-Zehnder Modulator

2.2.1 The Mach-Zehnder Interferometer Structure

Optical interferometry has played a crucial roll in development of physics, from Michelson and Morley's luminiferous aether experiment in 1887 [45, 46], to LIGO's ground-breaking gravitational wave detection in more recent years [47]. Regardless of its goal, an optical interferometer operates on a simple

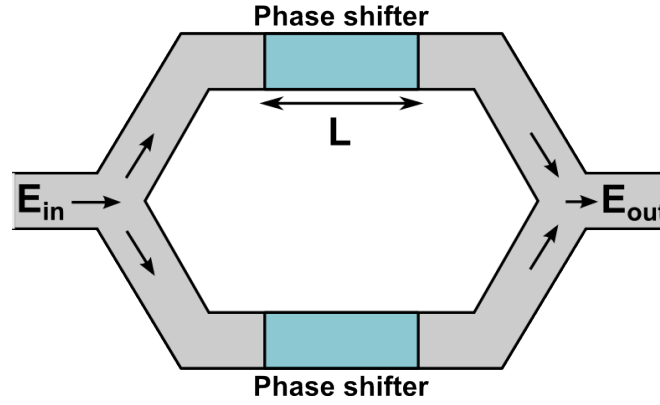


Figure 2.1: Basic structure of an MZM, with phase shifters in both arms.

principle: light waves can interfere with each other, combining constructively or destructively depending on their relative phase difference.

Intensity modulation in silicon is generally produced through phase shifters combined in an interferometric structure such as a ring resonator or Mach-Zehnder interferometer (MZI). As shown in Figure 2.1, an MZI essentially splits input light into two paths, applying a phase shift along one or both of the paths, before recombining it again at the output. Depending on the phase difference between the two paths, a change in the output intensity will be observed. In the case of the maximum π phase difference, total destructive interference occurs, and no light will be emitted. On the other hand, if the same phase shift is applied in both paths, total constructive interference occurs, and the output intensity will be at a maximum. An MZI is referred to as a Mach-Zehnder *modulator* (MZM) if the phase shift from either or both of the arms is modulated in order to produce a rapidly varying intensity at the output of the device. We will use the term MZM from here onwards.

Assuming that the MZM branches have an equal splitting ratio, that no light is lost coupling in and out of the MZM structure, and that no other aspects of the MZM introduce loss, the electric field of the outgoing light E_{out} can be related to the electric field of the incoming light E_{in} as

$$E_{out} = \frac{E_{in}}{2} \left(e^{i\Delta\phi_1} + e^{i\Delta\phi_2} \right), \quad (2.1)$$

where $\Delta\phi_1$ and $\Delta\phi_2$ are the respective phase shifts produced in the paths. Using the Euler identity, Equation 2.1 can be rewritten in terms of a cosine and

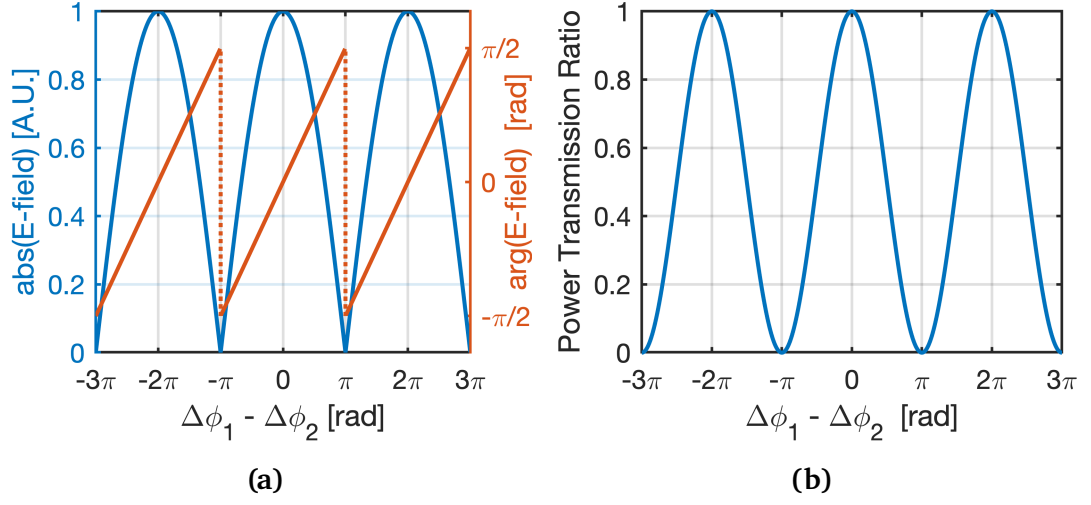


Figure 2.2: (a) Output electric field of an MZM, in terms of magnitude (abs) and phase (arg), against net phase variation in the arms. (b) Resulting output power of the MZM.

exponential as

$$E_{out} = E_{in} \cdot \underbrace{\exp\left(i \frac{\Delta\phi_1 + \Delta\phi_2}{2}\right)}_{\text{Phase variation}} \cdot \underbrace{\cos\left(\frac{\Delta\phi_1 - \Delta\phi_2}{2}\right)}_{\text{Amplitude variation}}. \quad (2.2)$$

The exponential term in Equation 2.1 represents the net phase variation of the output field, while the cosine term corresponds to the electric field's magnitude or amplitude. These are plotted in Figure 2.2(a), where the phase variation is wrapped at 90° . In direct-detection systems, as photodetectors are influenced only by the incident optical intensity or power, this phase variation is not detected. The output power of the MZM is obtained by squaring the magnitude of the electric field to produce

$$\begin{aligned} P_{out} &= |E_{out}|^2 = |E_{in}|^2 \cdot \cos^2\left(\frac{\Delta\phi_1 - \Delta\phi_2}{2}\right) \\ &= |E_{in}|^2 \cdot [1 + \cos(\Delta\phi_1 - \Delta\phi_2)]. \end{aligned} \quad (2.3)$$

This power varies as a sinusoid¹ as shown (normalised) in Figure 2.2(b), where the total phase difference between the paths, $\Delta\phi_1 - \Delta\phi_2$, is swept, illustrating peaks and nulls at even and odd integer multiples of π , respectively.

¹A sinusoid is any curve similar to the sine function but possibly shifted in phase, period and/or amplitude. Equation 2.3 gives two forms for the sinusoidal variation of MZM output power with phase. In this thesis, we will tend to use the $\cos^2(x)$ form for brevity.

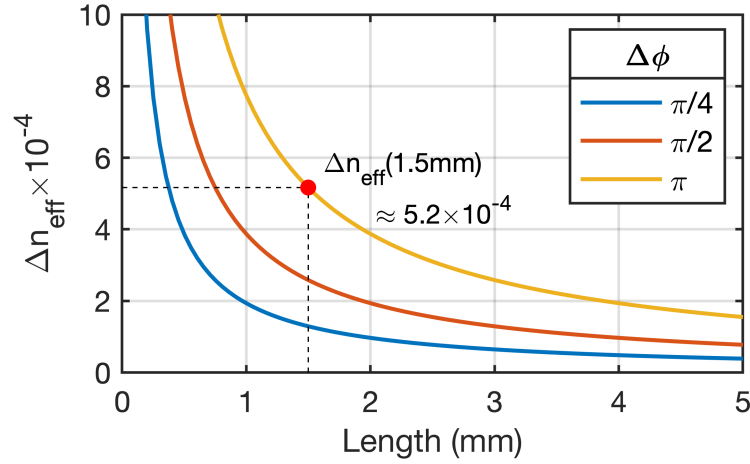


Figure 2.3: Required change in effective refractive index at 1550 nm to achieve phase shift of $\pi/4$, $\pi/2$ and π for various lengths.

For the purpose of optical communications, the phase shift in an MZM arm is achieved through a variation in the effective refractive index Δn_{eff} of the medium in which the light travels. This modifies the phase constant β of the optical wave, leading to an overall phase shift that depends on the light's free-space wavelength λ and the length of the phase shifting medium L according to

$$\Delta\phi = \beta L = \frac{2\pi}{\lambda} \Delta n_{eff} L. \quad (2.4)$$

The length L of the phase shifter in the MZM is often referred to as the active length or simply the length of the MZM (although the true length of the MZM structure will be noticeably longer). To be clear, in this thesis, any mention of the length of an MZM can be taken as referring to the phase shifter length. Figure 2.3 plots the refractive index change required to achieve target phase shifts at 1550 nm in devices of various lengths. Relating this to the experimental devices that will be used in this thesis, we note that to obtain complete destructive interference (a phase shift of π) using an MZM of length 1.5 mm, the Δn_{eff} between the arms must be approximately 5.2×10^{-4} . It should be noted that it is only the magnitude of this number that matters; whether the phase shift is positive or negative is irrelevant. It is also independent of the phase shifting technology, depending only on the device length and wavelength of operation.

Although it looks simple, Equation 2.4 underpins MZM design, as it demonstrates the inverse relationship between the required Δn_{eff} and L in phase shifters. As variations in the refractive index are typically controlled by (and proportional to) an applied voltage, the choice of device length is crucial. Longer lengths reduce

the voltage needed to achieve appreciable phase shift with a given technology, but there are practical limits to device lengths. With increasing length comes greater optical attenuation and reduced modulation bandwidth, presenting a challenge in device design. Bearing in mind that electronic driving circuits can have strict voltage limitations depending on the chosen technology and intended application, the phase shifter length, optical attenuation, and bandwidth must all be carefully optimised and balanced.

2.2.2 Silicon Waveguides

The paths used to carry the light in Figure 2.1 are silicon waveguides. They confine and guide light through internal reflection similarly to optical fibre, using a core of higher refractive index surrounded by a cladding of lower refractive index. However, the cross sections of silicon waveguides and optical fibre are different; the fabrication and processing steps used in PIC manufacture, such as etching and deposition, lend themselves to creating waveguides with rectangular geometry rather than circular. The core is silicon (Si), with a refractive index of 3.48 in the C-band at room temperature. The bottom layer of cladding on which the silicon waveguide sits is silicon dioxide (SiO_2), an insulating oxide of silicon with a refractive index of 1.44. The top layer of cladding, which surrounds the sides and top of the waveguide, can also be SiO_2 or left simply as air.

There are two main types of waveguide structure used in silicon photonic devices: the strip waveguide (also known as the channel, photonic wire, or ridge waveguide), and the rib waveguide (confusingly, also known as the strip-loaded ridge waveguide) [43]. The strip waveguide, shown in Figure 2.4(a), has a core that is surrounded entirely by cladding. It can provide tight bend radii and so is typically used for routing. Rib waveguides, as shown in Figure 2.4(b), have an added extra ridge of core material on either side of the central part. These protrusions allow electrical connections to be made to the waveguide and so are typically used for electro-optic devices such as MZMs.

The combination of the waveguide geometry and the refractive index of the core and cladding materials determines the effective refractive index of the waveguide n_{eff} (typically around 2.5) and the confinement of the optical modes that are travelling. Solving this can be done through finite-difference time-domain (FDTD), beam-propagation, or mode-solving techniques. In this work, a finite-element method (FEM) approach to mode solving was used with the commercial software *FemSIM* by Synopsys [48]. The resulting intensity of the

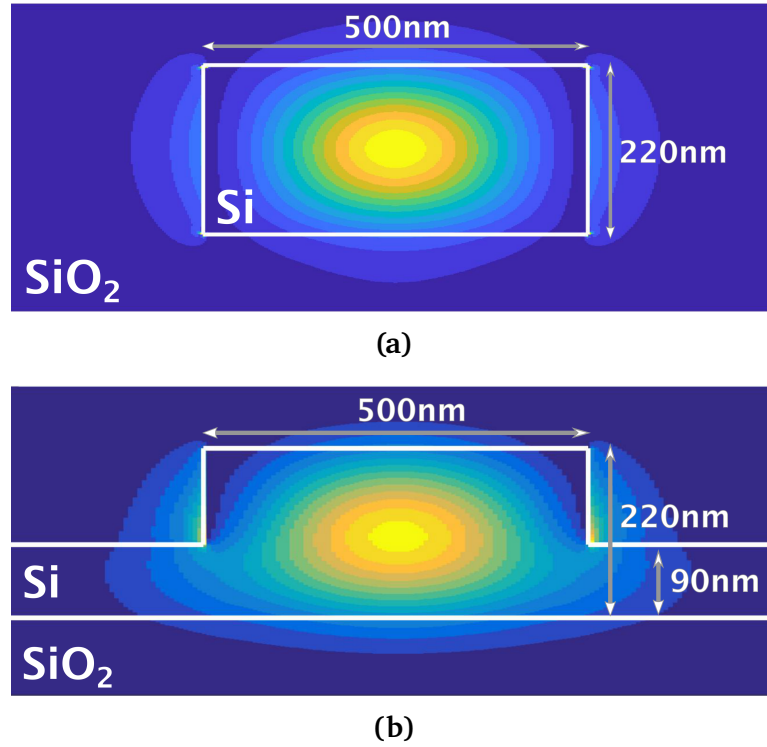


Figure 2.4: Simulated intensity profile of the optical TE mode at 1550 nm in (a) SiP strip waveguide, and (b) SiP rib waveguide.

TE mode profiles for the strip and rib waveguides are shown in Figure 2.4 as a heat map, demonstrating the strong confinement of the light to the core. The respective real parts of the effective refractive indexes obtained from the simulations for the strip and rib waveguides are 2.44 and 2.57. The attenuation coefficient used for the mode loss is calculated from the imaginary part of the refractive index through $\alpha = 4\pi\text{Im}(n_{\text{eff}})/\lambda$ [49]. These losses are generally outweighed by the other sources in the waveguides. Silicon waveguides can have high scattering loss due to side-wall roughness introduced during manufacture [50], which is typically in the region of 2 dB/cm to 3 dB/cm [43], although roughness reduction techniques during fabrication can reduce this loss to around 0.8 dB/cm [51]. Additional sources of loss in the waveguides include absorption due to metal in proximity, surface state absorption, and material absorption from any added dopant elements (which is discussed in Section 2.3.2) [43].

Silicon waveguides are a dispersive element, in that their effective refractive index is dependent on the wavelength of the propagating light. Therefore, the index of interest for certain analysis is often the optical group index of the waveguide, defined as

$$n_g = n_{\text{eff}} - \lambda \frac{dn_{\text{eff}}}{d\lambda}. \quad (2.5)$$

Looking simply at bulk silicon, a basic model using a first-order dependence of $-7.6 \times 10^{-5} \text{ nm}^{-1}$ and a more advanced Lorentz model with typical values inserted [43] both show a refractive index variation of less than 0.003 across the C-band. However, silicon waveguides exhibit a much stronger index-wavelength dependence than bulk silicon due to their geometry, and so their optical group index can differ significantly from their effective index [52, 53]. By simulating n_{eff} at various wavelengths and extracting the slope, $dn_{\text{eff}}/d\lambda$ is determined to be $-8.19 \times 10^{-5} \text{ m}^{-1}$ for the rib waveguide. This gives an n_g of 3.84 at 1550 nm. This value of optical group index will be used for the optical transit time considerations in section Section 4.5.3 and for the discussion of free spectral range later in this chapter.

2.2.3 Optical Fibre to Silicon Waveguide Coupling

The core diameter of standard single-mode fibre is $9 \mu\text{m}$, 18 times larger than the width of the silicon waveguide cores in Figure 2.4. This presents challenges when attempting to transfer light efficiently one to the other. Without any intermediate coupling device between the fibre and the waveguide, large optical power losses will be present from optical mode mismatch, refractive index differences and optical misalignment. An ideal coupler would losslessly transform the optical mode in the fibre to match the waveguide mode, prevent any back reflections at the fibre-waveguide interface, be insensitive to the polarisation and wavelength range of the light, and be implementable in a high-density wafer environment with large alignment tolerance.

Two main methods of coupling exist. Edge coupling (also known as in-plane or butt coupling) involves coupling the light via facets at the sides of the silicon chip. Edge couplers typically have high coupling efficiency (with correct alignment), almost flat wavelength dependence, and no polarisation-dependent loss. However, the polishing and etching that is normally required to form the facets do not lend themselves to large-volume fabrication. Alternatively, vertical coupling methods can be used, where the light is transmitted from the top surface of the chip via grating couplers—periodic structures that diffract the light into the desired path as shown in Figure 2.5(b). Typically, the optimal coupling angle between the fibre and waveguide is not fully perpendicular; rather, it is around 10° from the vertical to minimise strong back reflection due to second-order diffraction [54]. Grating couplers are convenient to fabricate, have relatively relaxed alignment requirements, have relatively wide optical

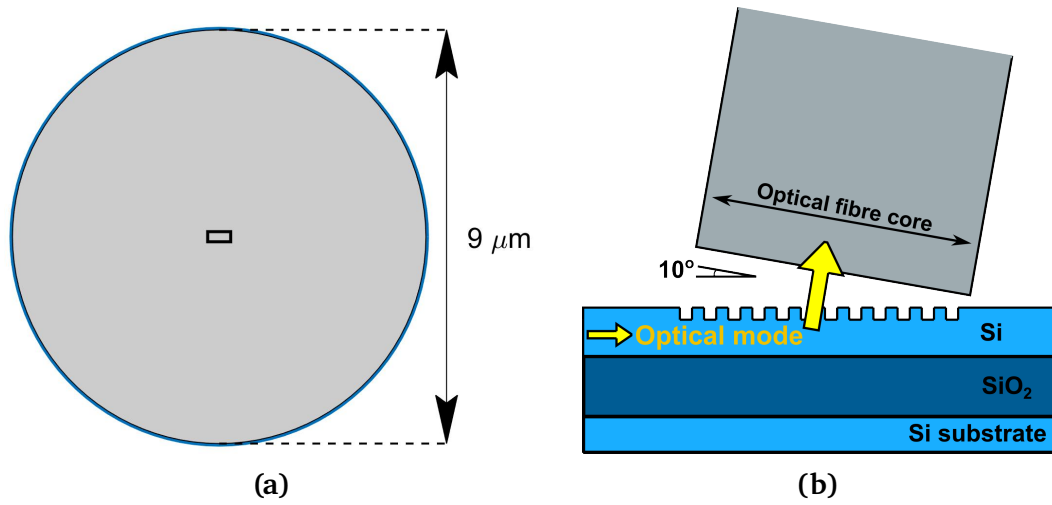


Figure 2.5: (a) Scale cross-section comparison of single-mode fibre core and typical silicon waveguide. (b) Not-to-scale profile of a 1D silicon grating coupler as an out-coupling device from a silicon waveguide.

bandwidths, and allow multi-point wafer characterisations as they can be placed anywhere along a waveguide. However, the simplest one-dimensional (1D) designs of grating couplers are intrinsically polarisation and wavelength sensitive, and so require careful optimisation to limit these effects. The SiP MZMs in this thesis use such grating couplers, and thus require tight polarisation control at their input in order to minimise their overall insertion loss. Alternatively, polarisation insensitive grating couplers can be formed, either through sub-wavelength grating features in a 1D design, or, more typically, by overlapping two orthogonal 1D gratings to form a two-dimensional (2D) design, where light is coupled into two paths depending on its polarisation. These are particularly important for coupling to silicon photodetectors at the receiver side of optical links, as the polarisation will vary depending on the route taken from the transmitter. It should be noted that positioning a fibre for vertical coupling as in Figure 2.5(b) naturally adds to the overall height of a SiP device. To avoid this unwanted increase in volume, planar packaging solutions have been created for grating couplers using total internal reflection at polished end facets [55]. Such a solution is used in this work for the MZMs studied in Chapter 4.

2.2.4 Modulation with an MZM

For electro-optic modulators, the phase shift in each path or arm of an MZM is achieved by controlling the refractive index with an applied voltage. The amount of voltage required for an appreciable phase shift depends on the choice of phase

shift technology used in the MZM; in general, for ease of circuit design and power consumption, smaller voltages are favoured. The metric used to evaluate the drive voltage requirements of an MZM is the half-wave voltage, or (V_π): the voltage needed to achieve a π phase shift in one arm of the MZM. In addition to total voltage needed, the chosen phase shift technology also influences the nature of the phase variation with voltage. For example, lithium niobate (LiNbO_3) phase shifters give a linear relationship between phase shift and applied voltage. In this case, if V_1 and V_2 are the voltages applied to the respective MZM arms, then their resulting phase shifts are $\Delta\phi_1 = \pi V_1/V_\pi$ and $\Delta\phi_2 = \pi V_2/V_\pi$, respectively, and the output electric field from Equation 2.2 can be expressed as

$$E_{out} = E_{in} \cdot \exp\left(i\pi \frac{V_1 + V_2}{2V_\pi}\right) \cos\left(\pi \frac{V_1 - V_2}{2V_\pi}\right). \quad (2.6)$$

The resulting output power in terms of the applied voltages is then

$$P_{out} = |E_{out}|^2 = |E_{in}|^2 \cdot \cos^2\left(\pi \frac{V_1 - V_2}{2V_\pi}\right). \quad (2.7)$$

In order to send data, the applied voltage must be varied or modulated. As an example, Figure 2.6 shows how the output intensity of the MZM varies when a sinusoidal voltage is applied to one of the arms. Two such voltages are illustrated in Figure 2.6(a): one with a peak-to-peak swing equal to V_π , and one with half that value. The other arm of the MZM has a constant voltage of $V_\pi/2$ applied in order to bias the modulator at its quadrature point—the midpoint of its output power.

Due to the sinusoidal dependence of the output power on drive voltage, the shape of the modulated optical signal is influenced by the amplitude of the modulating RF voltage. This shown in Figure 2.6(b), for continuous-wave light of 1 mW optical power passing into a lossless MZM with a V_π of 8 V, where sinusoidal RF drive signals with respective peak-to-peak voltages of 4 V and 8 V are applied to one arm while the other arm is held at quadrature. The smaller drive signal, shown in blue, mostly retains its sinusoidal shape when moving from the electrical to the optical domain as it is operating in the linear region of the MZM's transfer function. In comparison, the larger voltage, shown in orange, is distorted and compressed towards the top and bottom of the sinusoidal transfer function.

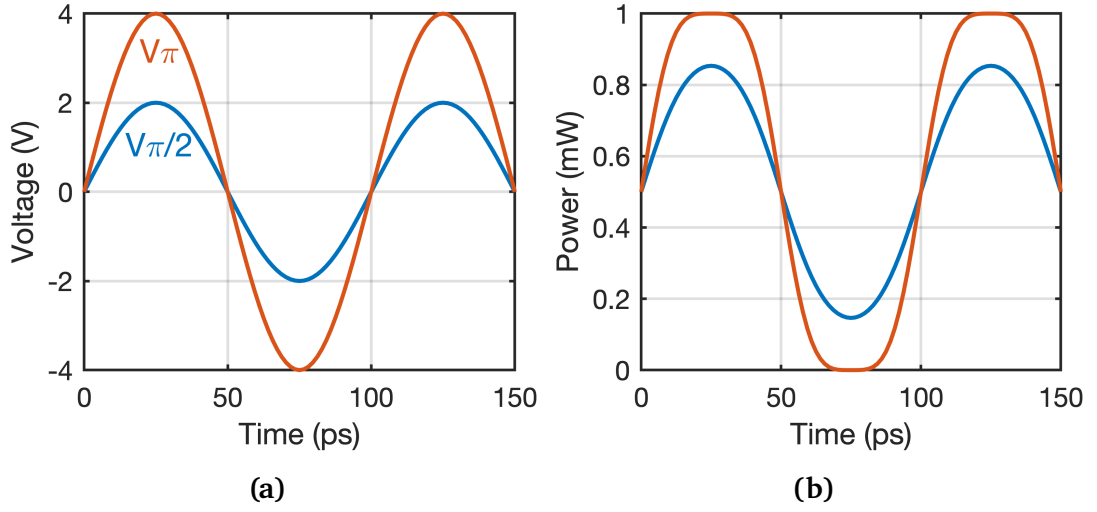


Figure 2.6: (a) 10 GHz sinusoidal drive voltages applied at quadrature to one arm of an MZM with a V_π of 8 V. (b) Corresponding modulated MZM output power against time.

2.2.5 MZM Drive Formats

There are two main drive formats that can be implemented with a standard MZM configuration, as shown in Figure 2.7 with simplified drive circuitry. The most basic form is the single-drive MZM, where only one arm receives a modulating RF voltage. The DC bias voltage can be applied to the other arm of the MZM, or to same arm that is being modulated by using a bias tee to isolate the DC and RF sources from one another. To achieve the maximum extinction ratio in the optical signal, the peak-to-peak RF voltage needs to be equal to V_π . The other drive format is the dual-drive, push-pull MZM, where both arms are driven with RF voltages of equal amplitude and opposite sign. In this case, assuming a linear relationship between voltage and n_{eff} (and, thus, phase shift), only half as much RF voltage is required in each arm in comparison with the single-drive format, as the arms “push and pull” the phase in opposite directions. As the power consumption in an MZM arm is proportional to the square of the RF voltage², this halving of the drive voltages actually halves the overall dynamic power consumption of the MZM compared with the single-drive design (assuming as an approximation that the additional driver circuitry for the second arm does not add any overhead to the power consumption).

There is another benefit to using a dual-drive design: if we again assume a linear relationship between voltage and n_{eff} , the net phase variation of the output

²Power consumption in MZMs will be discussed in more detail in Section 2.5.

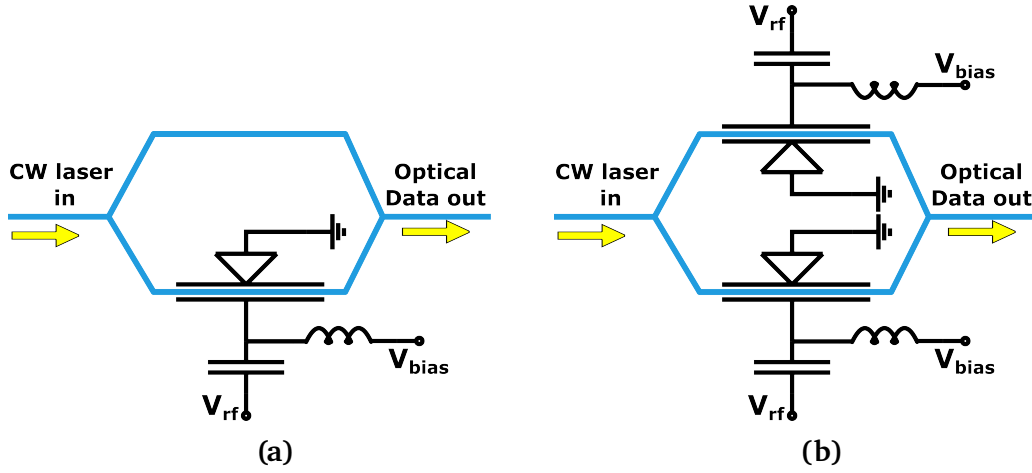


Figure 2.7: Simplified MZM driving circuitry showing (a) single-drive and (b) dual-drive schemes. The single-drive scheme applies RF voltage to only one arm, while the dual-drive scheme applies RF voltage to the second arm also.

electric field from Equation 2.6 is cancelled due to the push-pull nature of the applied voltages (i.e., $V_2 = -V_1$, leading to the exponential term in the equation going to zero). This is important when considering signal transmission over relatively long distances of fibre. A signal is said to have “chirp” if there is a time variation in its instantaneous frequency $f(t)$. Instantaneous frequency is itself related to the time derivative of instantaneous phase $\phi(t)$ through $f(t) = \frac{1}{2\pi} \frac{d\phi}{dt}$. Hence, any residual net phase variation in the output electric field of a modulated MZM will lead to a time-varying frequency and chirp. Although direct-detection systems do not detect variations in phase, the interaction of chirp with chromatic dispersion over kilometres of fibre leads to pulse broadening or compression which can affect signal quality. This total cancellation of chirp only happens if the phase shifters in the dual-drive MZM change only the phase of the optical field in their path (not the amplitude), and do so in a purely linear fashion. Neither of these requirements are fulfilled with SiP depletion-mode phase shifters, as will be shown in Section 3.2.2. However, SiP MZMs have been shown to have a small, negative chirp for certain bias conditions, which can actually be beneficial in counteracting the pulse broadening effect over small distances of fibre [56].

There is a third drive format for MZMs, known as the series push-pull (SPP) or single-drive push-pull MZM [57]. In this design, both arms are connected as back-to-back diodes with a shared DC bias applied via their common node via a long, on-chip inductive trace. This design leads to lower microwave losses in the arms, resulting in potentially higher bandwidths, although it does require

a full V_π to be applied for maximum extinction. SPP-MZMs must be fabricated differently to dual-drive MZMs to create the back-to-back design of the phase shifters, and so a dual-drive MZM cannot be operated in series push-pull fashion. No SPP-MZM devices were available for experiments during this work, so they are not discussed further. A thorough description of the design and testing of an SPP-MZM is available in [58].

2.2.6 Unbalanced MZMs

In order to operate a single-drive MZM at quadrature, the bias voltage needs to be set to $V_\pi/2$. Achieving this value through capacitive coupling on a driver chip can be difficult for the relatively high V_π of SiP MZMs. The breakdown voltage of any capacitor scales with dielectric thickness, presenting a challenge as CMOS drivers move towards deep sub-micron technology nodes. For example, in 65 nm CMOS, which was used for the switched-capacitor driver described in Section 3.4.1, the breakdown voltage of on-chip capacitors is approximately 5 V [59, 60], which may leave little room for the RF voltage to cycle around the required bias. For a LiNbO₃ MZM, where both positive and negative bias voltages may be applied, a push-pull, dual-drive design allows the bias to be divided between each arm and thus halved to $\pm V_\pi/4$. However, this is not possible with SiP depletion-mode MZMs, which (as the name suggests) must keep their PN-diode phase shifters in depletion with reverse bias. In addition, as will be shown in Section 3.3.2, the modulation efficiency of SiP phase shifters depends on the applied bias voltage; for equal modulation in each arm, it is best to keep the two bias voltages equal.

Alternative methods of reaching the MZM's quadrature point include adjusting the wavelength of the light used, or adding thermal phase shifters to the arms. We will first address the wavelength tuning method. To do so requires an unbalanced or asymmetric MZM, where a small length difference ΔL is introduced between the two optical paths. The added distance that the light must travel in the longer arm brings about a natural, extra phase shift, resulting in a periodic variation of P_{out} with wavelength. This period, defined as the free spectral range (FSR), can be expressed as [61]

$$FSR = \frac{\lambda^2}{n_g \Delta L}, \quad (2.8)$$

where n_g is the optical group index of the waveguide. The FSR is plotted against ΔL in Figure 2.8(a), and is shown to decrease quickly as the imbalance in arm length increases. It should also be noted that the FSR varies little across the the

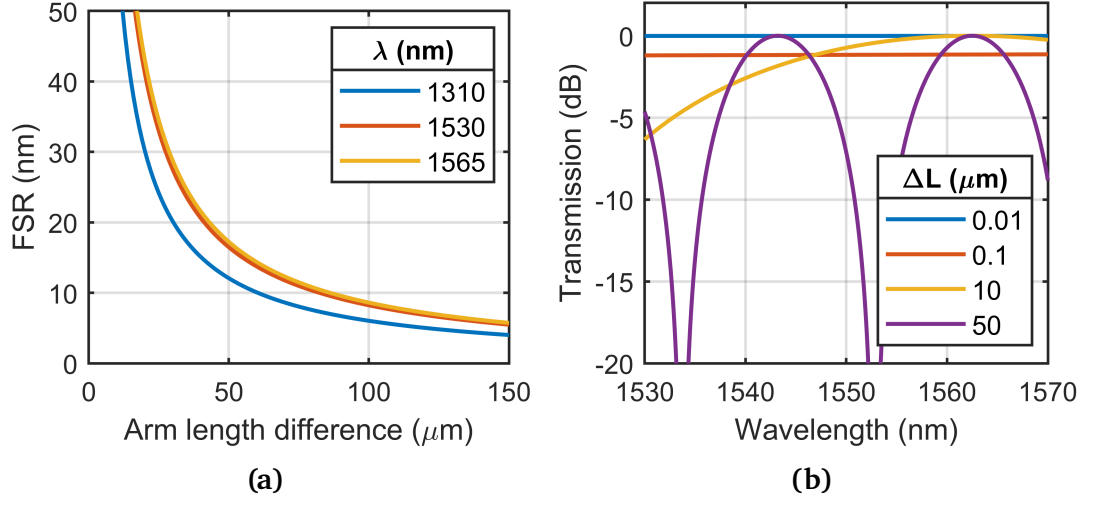


Figure 2.8: (a) MZM FSR variation against length imbalance in the C-band and O-band. (b) MZM transmission spectrum for various length imbalances. The FSR is the distance between the peaks in the spectrum.

range of wavelength in the C-band (1530 nm to 1565 nm), but does change by approximately 30% when the wavelength chosen is in the O-band (1310 nm), indicating that the choice of length imbalance should be chosen appropriately for the operating band. The dependence of the MZM spectrum on the length imbalance follows simply from Equations 2.3 and 2.4, where the arm lengths used for $\Delta\phi_1$ and $\Delta\phi_2$ are L and $L + \Delta L$, respectively, to give a sinusoidal relationship between output power and wavelength:

$$P_{out} = |E_{in}^2| \cos^2 \left(\frac{\pi}{\lambda} n_{eff} \Delta L \right). \quad (2.9)$$

This sinusoidal dependence is shown in Figure 2.8(b), for various length imbalances³. As expected, the MZM is optically broadband when no imbalance is introduced. However, even a small imbalance such as 50 μm causes a sufficiently small FSR to provide transmission variation around the wavelengths of interest. By designing an MZM with suitable FSR, the laser wavelength can be adjusted to operate the MZM at quadrature. This is highly advantageous for device testing in an experimental setting, as it allows the operation point of the MZM to be quickly adjusted or swept without disturbing the bias or temperature of the device. If deployed in a data centre setting, however, where optically broadband devices may be required for the purposes of wavelength-division multiplexing, it may be preferable to use balanced MZMs and tune the operation

³The transmission spectrum is typically plotted in decibels rather than a linear scale, hence the difference in appearance of our sinusoid.

point through additional thermal or DC phase shifters in the arms.

It should be noted that if too small a length imbalance such as 0.1 μm is used, then quadrature points may fall far outside the operating wavelength range of the laser. In this case, the MZM remains optically broadband, but can exhibit an additional insertion loss if the maximum transmission point also occurs at a distant wavelength. This can be problematic, as such an imbalance can be introduced via fabrication errors or variations during MZM creation. These variations include SOI wafer thickness, photolithography mask accuracy and density, and plasma density [62]. Variations can occur over time from wafer to wafer or batch to batch, but can also happen on a single run within the same wafer. For example, in a 193 nm lithography process, even normally identical SiP MZMs that are closely spaced on the same die have been shown to have average intra-die wavelength shifts of 0.2 nm, while those on different dice show average inter-die wavelength shifts of up to 1.73 nm [62]. This highlights the need for additional phase shifters in MZMs that must be operated at specific wavelength. It should be noted that future extreme-ultraviolet lithographic processes at 13.5 nm, if used eventually for SiP devices, are expected to improve this spectral accuracy [62] and may reduce the need for such control.

2.2.7 Thermo-optic Effect in Silicon

Bulk silicon exhibits quite a large, linear, variation of refractive index with temperature [63, 42], as shown by its thermo-optic coefficient, given here at room temperature:

$$\frac{dn}{dT} = 1.86 \times 10^{-4} \text{ K}^{-1} \text{ at } 300 \text{ K.} \quad (2.10)$$

Given the refractive index of silicon at a reference temperature T_0 , this coefficient can be used to calculate resulting refractive index at a temperature T_1 according to $n_{\text{eff}}(T_1) = n_{\text{eff}}(T_0) + \frac{dn}{dT}(T_1 - T_0)$. Such a large variation of refractive index with temperature facilitates the use of small thermal phase shifters in silicon MZM arms. A quick calculation using Equation 2.4 shows that a π phase shift can be achieved in a 500 μm length of silicon with just 8.3 $^\circ\text{C}$ change in temperature. Although the bandwidth of thermal phase shifters is too low to be used for high-speed modulation, they can be used for convenient biasing of the MZM [64]. Unfortunately, this large thermo-optic effect also presents a challenge in data centres, where uncooled transceivers are typically expected to operate with case temperatures in a range of 0 $^\circ\text{C}$ to 70 $^\circ\text{C}$, which can be extended to -5 $^\circ\text{C}$ to 85 $^\circ\text{C}$ in certain specifications [64, 65]. It is commonly known that the behaviour of SiP

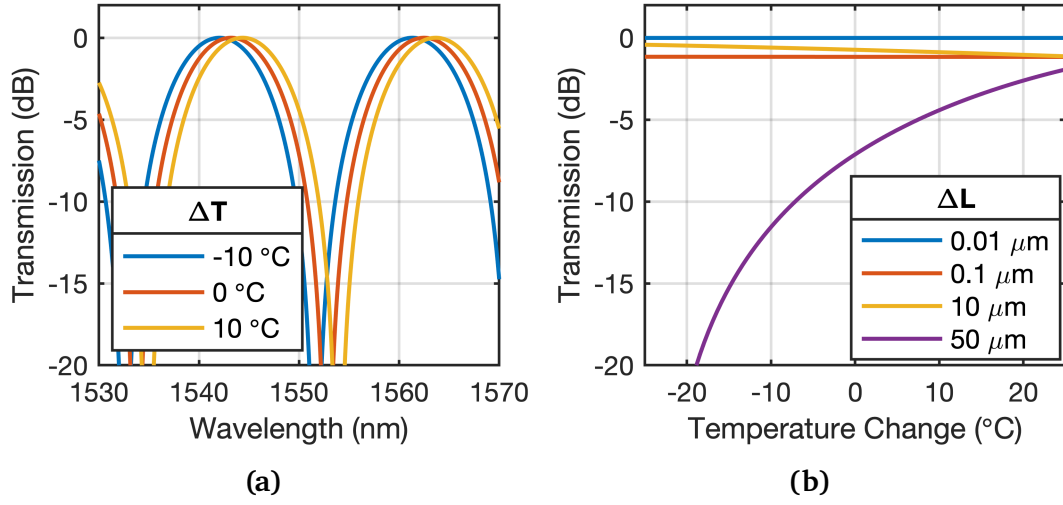


Figure 2.9: (a) Transmission spectrum of an MZM with 50-μm length imbalance for various global temperature changes. (b) Variation in transmission with temperature change for various length imbalances.

ring resonators is strongly temperature dependent, placing strict requirements on their temperature control. However, temperature variations can also affect unbalanced and, in certain cases, balanced MZMs, as shown below.

For an unbalanced MZM, the variation of its spectrum with an overall⁴ temperature change of ΔT is given again by Equation 2.9, where the refractive index now has a temperature dependence $n_{\text{eff}}(T)$ that follows Equation 2.10 to give:

$$P_{\text{out}} = |E_{\text{in}}^2| \cos^2 \left(\frac{\pi}{\lambda} n_{\text{eff}}(\Delta T) \Delta L \right). \quad (2.11)$$

As shown in Figure 2.9(a), an overall variation of 10 °C has a noticeable impact on the spectrum of an MZM with a 50 μm imbalance, shifting it by 0.8 nm. Figure 2.9(b) shows this dependence of the shift in transmission power with overall temperature change for various length imbalances. As could be expected, very small imbalances such as 0.01 μm and 0.1 μm are unaffected, but longer imbalances (that would be chosen to reach quadrature via wavelength tuning) can have a strong dependence of transmission power with temperature. This implies that unbalanced MZMs may require temperature control to keep them at a chosen bias point when used in data centre settings, either through external metal heaters [66] or silicon thermal phase shifters [67], although there are methods of minimising this temperature sensitivity. CMOS-compatible passive compensation of temperature variations in unbalanced MZMs has

⁴By overall, we mean a homogenous temperature change throughout the whole MZM.

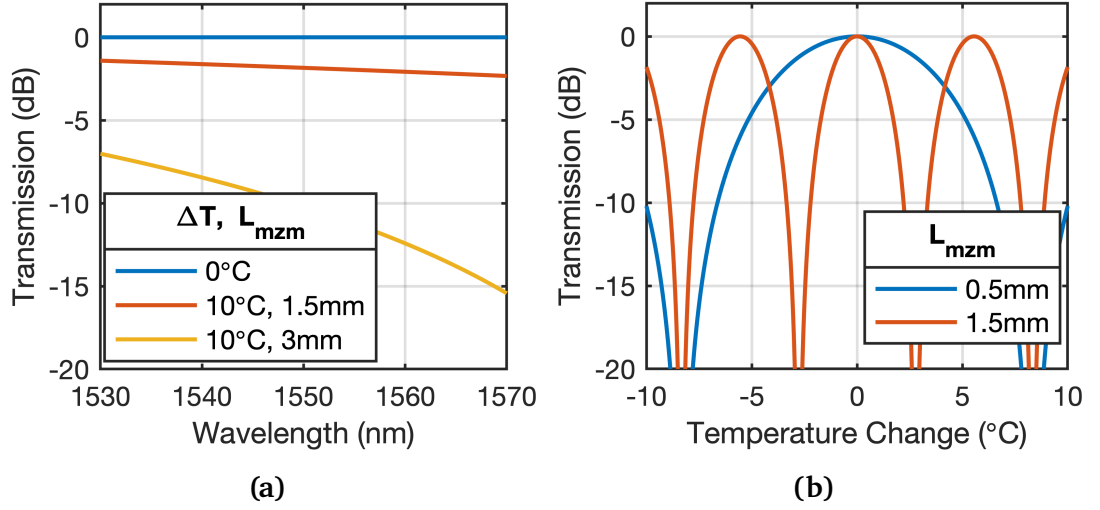


Figure 2.10: (a) Balanced MZM transmission spectrum (b) Balanced MZM power transmission against arm temperature difference at 1550 nm.

been demonstrated by introducing a deliberate difference in the widths of the waveguides in each arm and carefully optimising the arm lengths[68].

Although a balanced MZM is unaffected by an overall temperature change, it can still be affected by localised temperature difference between its arms. This can be introduced deliberately, for the purposes of thermal phase shifting as discussed above, or in an undesired fashion if different drive voltages are applied to the respective MZM arms. To see the effect of this temperature difference, similarly to the derivation for the unbalanced MZM spectrum, we start from Equations 2.3 and 2.4 but add a temperature difference ΔT in the arms rather than a length imbalance. This leads to the following equation for dependence of a balanced MZM's output power on temperature difference between the arms:

$$P_{out} = |E_{in}^2| \cos^2 \left(\frac{\pi}{\lambda} \Delta n_{eff}(\Delta T) L \right). \quad (2.12)$$

We note from the presence of λ in this equation that an FSR has actually been introduced to the spectrum due to this temperature imbalance in the arms. This wavelength dependence is quite flat for relatively short MZMs, but can be significant for longer MZMs, as shown in Figure 2.10(a) for a 10 °C imbalance in 1.5 mm and 3 mm MZMs operating at 1550 nm. Of more concern, however, is how the transmission of a single wavelength varies with temperature. This depends strongly on device length, as shown in Figure 2.10(b). Even for relatively short lengths of 0.5 mm and 1.5 mm, we note that just a few degrees of temperature difference in the arms can move the MZM away from its desired

operating point. However, in practice, simply allowing a settling time for the arms to reach a stable temperature differential and operating condition solves this problem. In addition, the SiP MZMs used in this thesis are typically operated with the same average drive voltage in each arm, which minimises any voltage-induced temperature difference.

2.3 Modulation in Silicon

2.3.1 Complex Refractive Index and Kramers-Kronig

The complex refractive index of any material can be written as

$$\bar{n} = n + i\kappa, \quad (2.13)$$

where n is the refractive index and κ is the optical extinction coefficient. The real part n in this definition relates, in the usual sense, the phase velocities of light in vacuum and in the medium in question. The imaginary part κ determines the optical absorption of the medium, and is related to the linear absorption coefficient α by

$$\kappa = \frac{\alpha\lambda}{4\pi}. \quad (2.14)$$

where λ is the wavelength of the propagating light.

Optical modulators are generally⁵ based on a shift in the real or imaginary refractive index of the material in which the optical field propagates. Changes in n produce phase shifts, which can then be converted to changes in amplitude using MZMs or MRRs as discussed previously, while changes in κ can be used directly in electroabsorption modulators. The well-known Kramers-Kronig relation, which links the real and imaginary part of any analytic complex function, can be used to calculate the real part of the refractive index from the imaginary part (and vice versa). More crucially, it also holds for *changes* in the real and imaginary parts of the complex index. It is shown here relating changes in the refractive index to changes in the linear absorption coefficient [33]:

$$\Delta n(\omega) = \frac{c}{\pi} \text{PV} \int_0^\infty \frac{\Delta \alpha(\omega')}{\omega'^2 - \omega^2} d\omega', \quad (2.15)$$

where ω is the angular carrier frequency of the light, c is the speed of light in vacuum, and PV indicates that the Cauchy principal value of the integral must

⁵Notable exceptions include spatial light modulators and acousto-optic modulators.

be used due to the singularity at $\omega' = \omega$. This relation is particularly useful for calculating the refractive index change in a material due to an applied stimulus, as it is much easier to make accurate measurements of the optical absorption of a material than its refractive index. The Kramers-Kronig relation also demonstrates that it is impossible to achieve a change in refractive index without a corresponding change in absorption, which can be detrimental to optical transmission performance through changing insertion loss and the creation of chirp. The strength of the variation of α with n depends on the material in question. LiNbO_3 , for example, has a particularly low absorption coefficient ($<10^{-2} \text{ cm}^{-1}$) in the telecom wavelengths [69], and any active changes in α resulting from LiNbO_3 phase shifters are insignificant in terms of the overall device operation. This is not necessarily the case with silicon, as we will see in Section 3.2.2.

2.3.2 Achieving Phase Shift in Silicon

Variation of the refractive index of a material with an applied voltage can be achieved directly through the electro-optic effect or via changes in the free carrier concentration. The linear electro-optic effect, or Pockels effect, as used in lithium niobate phase shifters, requires non-centrosymmetric crystals⁶ and therefore is not present in silicon, which has centrosymmetry unless the crystal is strained by the introduction of other materials such as silicon nitride [71]. The quadratic electro-optic effect, or Kerr effect, is present in all materials, although its effect in silicon is relatively weak. To illustrate this, we will investigate the required electric field strength that must be applied to achieve the necessary Δn of 5.2×10^{-4} for a full π phase shift in a 1.5 mm long MZM, as was shown in Figure 2.3. Electric field strengths in this context are traditionally shown with units of V cm^{-1} . As voltages in silicon phase shifters are typically applied across rib waveguides with contact distances in the order of $1 \mu\text{m}$ to $5 \mu\text{m}$, we will make the numbers more readily relatable by listing in brackets the electric field strength with units of $\text{V } \mu\text{m}^{-1}$ also.

A formula for the relevant refractive index change due to the Kerr effect at a given frequency of light can be derived using an anharmonic oscillator model, where the bound electrons are displaced from their equilibrium positions under

⁶A crystal has centrosymmetry, or inversion symmetry, if for each of its points (x, y, z) in its point group there is an indistinguishable point $(-x, -y, -z)$ [70], i.e., it is symmetric when reflected through its centre point.

the influence of an applied electric field E , to give [72, 34]

$$\Delta n = -\frac{3q_e^2(n^2 - 1)E^2}{2nm^2\omega_0^2(\omega_0^2 - \omega^2)x^2}. \quad (2.16)$$

Here, q_e is the charge on the electron, n is the unperturbed refractive index, m is the free electron mass, ω is the angular frequency of the light, ω_0 is the resonant frequency of the oscillator, and x is the average oscillator displacement. This resonant frequency occurs close to the peak of the fundamental absorption band, which for a semiconductor is typically in the region of 4 eV [73, 74] or 1 PHz, and thus ω_0 can be assumed to be $2\pi \times 10^{15} \text{ rad s}^{-1}$. This is far higher than the optical frequencies of telecom bands, so Equation 2.16 is actually wavelength independent in our region of interest. The value for x can be expected to be roughly equal to the covalent radius of the semiconductor component atoms, which for silicon is approximately 0.1 nm [72]. Inserting the appropriate values into this formula shows that an applied electric field of in the order of 10^6 V cm^{-1} ($1000 \text{ V } \mu\text{m}^{-1}$) is required to achieve the necessary Δn of 5.2×10^{-4} . Given that the dielectric breakdown strength of intrinsic crystalline silicon is in the order of 10^5 V cm^{-1} ($100 \text{ V } \mu\text{m}^{-1}$) [73], this is not an ideal technique to use for index change. (The prospect of generating swings of 100 V or more with high-speed transceiver circuitry is obviously another prohibitive hurdle.)

The Franz-Keldysh effect is the change in the optical absorption of a semiconductor when an electric field is applied, and is caused by tunnelling of electrons and holes between valence and conduction band states. The electron and hole wavefunctions become Airy functions with characteristic tails that extend into the bandgap [75]. This change in absorption, known specifically as electroabsorption, is accompanied due to Equation 2.15 by an associated change in refractive index known as electrorefraction. Soref and Bennet [34] calculated the strength of electrorefraction in silicon from experimental measurements of electroabsorption over a range of wavelengths. Fig. 2 in their paper shows that Δn peaks at 1040 nm and 1090 nm, giving values of -1.7×10^{-5} and 1.2×10^{-5} , respectively, for the maximum applied field of 10^5 V cm^{-1} ($100 \text{ V } \mu\text{m}^{-1}$). Unfortunately, these values are still an order of magnitude less than the required Δn of 5.2×10^{-4} , whilst also lying outside the telecom bands. Things are even worse at 1550 nm, with the same field giving a Δn of approximately 1.5×10^{-6} .

As discussed in Section 2.2.7, the thermo-optic effect in silicon is strong enough

to provide suitable phase shifts in short lengths of waveguides. Optimised designs using air-gap trench isolation, silicon undercuts and the addition of additional materials such as titanium nitride can produce low-voltage, low-power thermal phase shifters [76]. However, they are not suitable for high-speed modulation as they have bandwidths in the order of kHz, and so instead find their niche in data centre switching applications.

Today, the most commonly used variation of refractive index for modulation in silicon is invoked by the free-carrier plasma dispersion effect, where light is scattered by mobile carriers (free electrons and holes). These free carriers are typically introduced to silicon through a semiconductor doping process with boron and phosphorous. A theoretical description of this free-carrier absorption and corresponding refractive index change follows the Drude-Lorentz equations [72, 34], relating index and absorption changes linearly to carrier concentrations as

$$\Delta n = -(q_e^2 \lambda^2 / 8\pi^2 c^2 \epsilon_0 n) [\Delta N_e / m_{ce}^* + \Delta N_h / m_{ch}^*], \quad (2.17)$$

$$\Delta \alpha = (q_e^3 \lambda^2 / 4\pi^2 c^3 \epsilon_0 n) [\Delta N_e / m_{ce}^{*2} \mu_e + \Delta N_h / m_{ch}^{*2} \mu_h], \quad (2.18)$$

where ΔN_e and ΔN_h are the respective concentrations of holes and electrons, q_e is the electronic charge, ϵ_0 is the permittivity of free space, n is the refractive index of unperturbed crystalline silicon, m_{ce}^* is the conductivity effective mass of electrons, m_{ch}^* is the conductivity effective mass of holes, μ_e is the electron mobility, and μ_h is the hole mobility. However, in reality, the measured absorption spectra and refractive index shift of silicon do not correspond exactly with the theoretical predictions of the Drude-Lorentz equations. To resolve this, Soref and Bennett [34] fitted a power-law model to aggregated measurements, resulting in widely used empirical equations for Δn and $\Delta \alpha$ that are shown here for a wavelength of 1550 nm:

$$\Delta n = -(8.7 \times 10^{-22} \Delta N_e + 8.5 \times 10^{-18} \Delta N_h^{0.8}), \quad (2.19)$$

$$\Delta \alpha = 8.5 \times 10^{-18} \Delta N_e + 6.0 \times 10^{-18} \Delta N_h, \quad (2.20)$$

where the changes in carrier concentration ΔN_e and ΔN_h have units of cm^{-3} . It should be noted that these equations are wavelength dependent; equivalent equations at 1310 nm were also reported in the same work, and, later, for wavelengths over the 1000 nm to 14 000 nm infrared range [77].

As was expected from the Kramers-Kronig relation, Equations 2.19 and 2.20 show

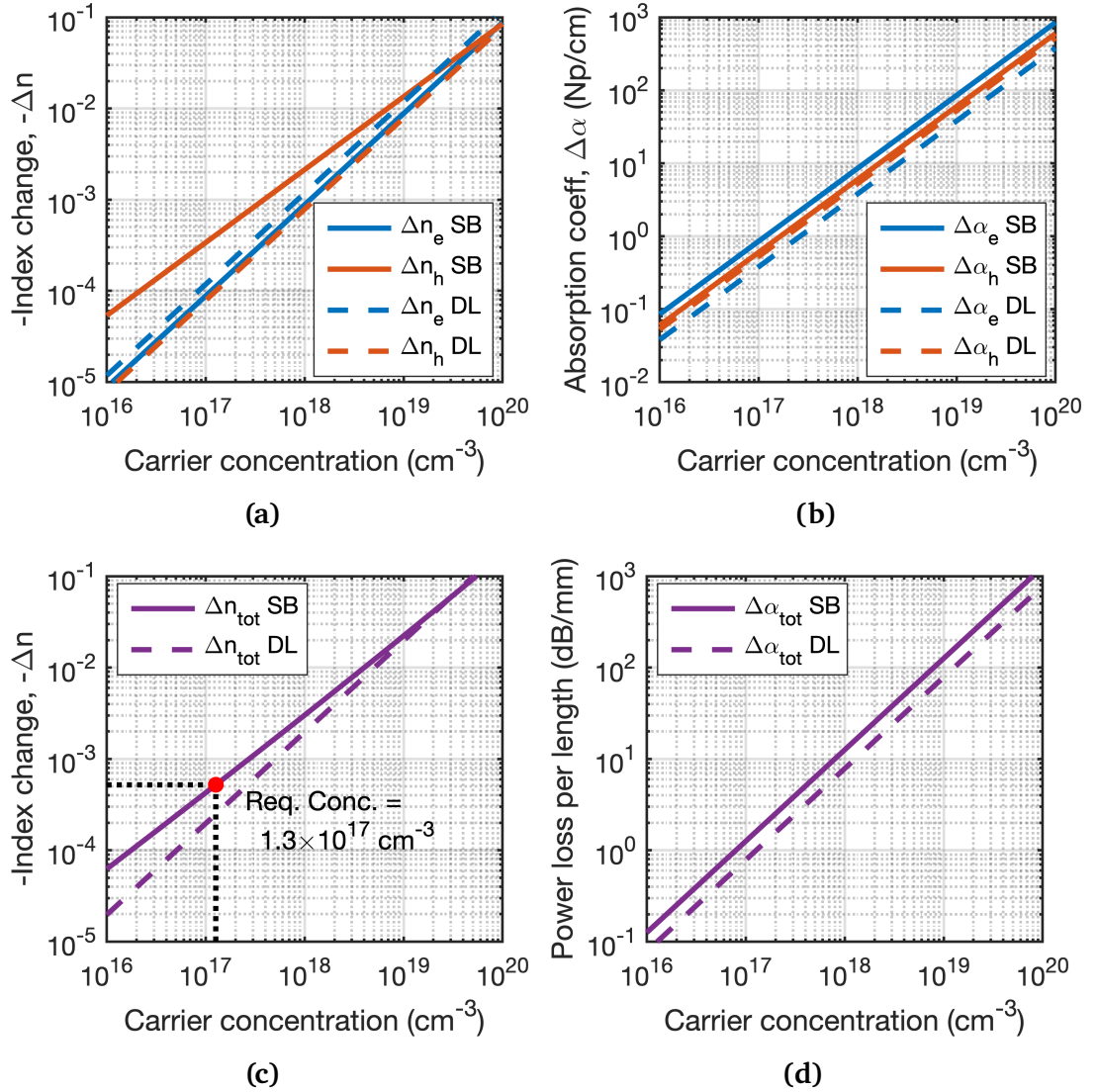


Figure 2.11: (a) Refractive index changes and (b) absorption coefficient changes due to concentrations of electrons and holes. (c) Total refractive index change from both electrons and holes. (d) Resulting optical power loss in dB/mm. SB: Soref-Bennett model. DL: Drude-Lorentz model.

that changes in the refractive index and absorption of silicon are linked. In fact, there is a fundamental trade-off between modulation efficiency (related to Δn) and optical loss (related to $\Delta \alpha$) in silicon. This is illustrated in Figure 2.11, which looks at how carrier concentrations affect the refractive index and absorption. For comparison, also shown are the corresponding index and absorption changes as calculated from the Drude-Lorentz model. Figures 2.11(a) and 2.11(b) show the changes in refractive index and absorption coefficient respectively against carrier concentration, where the contributions from the electrons and holes have been plotted separately for clarity. For ease of viewing in the plots, the negative

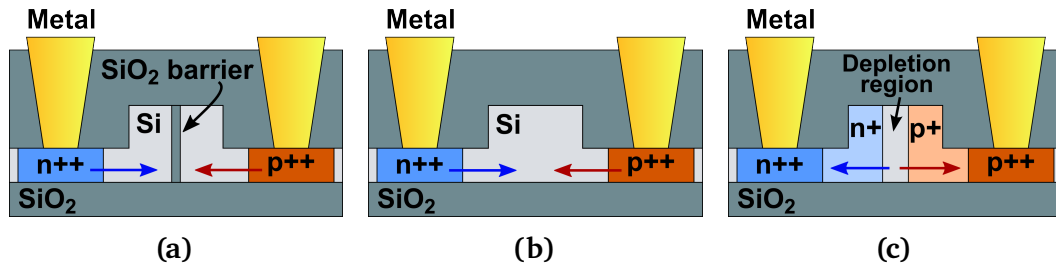


Figure 2.12: Cross sections of silicon phase shifters made from rib waveguides using (a) carrier accumulation design, (b) carrier injection design, and (c) carrier depletion design. The blue and red arrows indicate the movement of free electrons and holes respectively under the applied bias.

of the refractive index change ($-\Delta n$) is used. Noticeably, an increase in hole concentration gives larger index change than a corresponding increase in electron concentration, while also giving a smaller increase in absorption. This makes holes a more ideal free carrier than electrons for achieving phase shift in silicon; careful choice of the concentrations and distributions of holes and electrons in silicon phase shifters is therefore crucial to optimise performance [78]. To relate these carrier concentrations and index changes to a familiar number, Figure 2.11(c) plots the combined index change from equal concentrations of holes and electrons on a log scale. From this, the required concentration to achieve an index change of 5.2×10^{-4} is identified as $1.3 \times 10^{17} \text{ cm}^{-3}$. This gives a ball-park figure for the carrier concentrations that are required to be present at some point during modulation in the waveguides of silicon phase shifters. Whether the carriers are added or removed during the course of modulation does not matter; it matters only that the concentration *changes* by this amount. Finally, Figure 2.11(d) plots the optical loss per length resulting from the absorption coefficient in units of dB/mm. We note that this ball-park carrier concentration of $1.3 \times 10^{17} \text{ cm}^{-3}$ results in a loss of 0.5 dB/mm. This trade-off between phase shift and absorption will be discussed in more detail for silicon depletion-mode MZMs in Section 3.3.2. In reality, the required doping level that is suggested here is an underestimate as the optical mode in the waveguide is unevenly distributed, being concentrated more in the centre of the waveguide where the depletion region also resides.

2.3.3 Silicon Phase Shifter Structures

There are three main methods of manipulating the carrier concentration of a silicon waveguide [79]: carrier accumulation, carrier injection, and carrier

depletion. They are shown in Figure 2.12 and described below in their simplest forms. It should be noted that, in reality, more advanced designs vary in their layout and doping profile. In particular, the undoped silicon described here is often lightly doped in order to optimise modulation efficiency and bandwidth. In each case, a rib waveguide with heavily doped edges is used to facilitate electrical contact for the RF signal through metal vias, as shown by the $p++$ and $n++$ regions in the figures. Multiple of these vias are placed along the phase shifter length, linking the PN junction at many points to the electrode that runs above.

In carrier accumulation (Figure 2.12(a)), the centre of the waveguide is intrinsic, undoped Si, but a capacitive structure is formed by isolating the two halves of the waveguide from each other using a thin insulating layer of SiO_2 . A forward bias applied across the SiO_2 dielectric layer causes the accumulation of charges or free carriers at the Si- SiO_2 boundaries, thus varying the carrier concentration of the waveguide. For carrier injection (Figure 2.12(b)), the the centre of the waveguide is again undoped, intrinsic silicon. Applying a forward bias to the device causes free electrons and holes to be injected into the intrinsic region, varying the overall carrier concentration of the waveguide. More optimised carrier injection designs can use doped rather than intrinsic silicon in the central waveguide to form a PN or PIN junction. In carrier depletion designs (Figure 2.12(c)), the central part of the waveguide must be lightly doped to form p-type and n-type regions, resulting in a PN diode with a typical central depletion region in which no free carriers are present. The width of this depletion region widens with increased reverse bias, allowing the overall carrier concentration of the waveguide to be varied.

As will be discussed in Section 2.6, depletion-mode MZMs currently retain the crown for highest achievable modulation bandwidths in silicon, hence the SiP MZMs studied in this thesis will all use depletion-mode designs. Despite this, carrier injection and accumulation designs do offer superior modulation efficiency. Therefore, an ongoing challenge in SiP MZM design is to improve the efficiency of depletion-mode MZMs while maintaining high bandwidth.

2.4 Lumped vs. Travelling-Wave MZMs

As the V_π of an MZM reduces with length, it is favourable to use MZMs with long active regions to keep the required drive voltages low. However, as will be shown in Section 3.3.3, the RC bandwidth of an MZM in combination with its

associated driver reduces when the active length of the MZM increased. This leads to a balancing act between bandwidth and V_π in device design. The simplest description of the modulation bandwidth of an MZM is given by the RC time constant of the phase shifters in the arms. For this, the amplitude-frequency response can be considered in the usual way through the gain G_C across a capacitor C in series with a resistor R , given by the magnitude of the capacitor's complex frequency response $H_C(i\omega)$:

$$G_C = |H_C(i\omega)| = \frac{1}{\sqrt{1 + (\omega RC)^2}}, \quad (2.21)$$

where ω is the angular frequency of the stimulus voltage. If needed, parasitic resistive, capacitive, and inductive contributions from other elements on the chip such as bondpads or wire bonds can also be included. The 3 dB bandwidth of the MZM follows as

$$f_{3dB} = \frac{1}{2\pi RC}. \quad (2.22)$$

The above description of the bandwidth holds accuracy only in the case of the lumped-element regime, where the transit time of the RF signal through the electrodes of the phase shifters is sufficiently short. After a certain maximum length l_{LE} , which depends heavily on the frequency of the RF signal in question, the electrodes can no longer be considered as a simple lumped-element circuit which can be treated purely in terms of resistance, capacitance and inductance. Instead, the signal along the line must be considered in a distributed fashion, with its characteristics varying depending on its position on the electrode. In the case of a long MZM, this means the frequency response will be described and limited by three main effects [80]: the attenuation of the RF signal as it travels along the electrode, the velocity mismatch between the optical and electrical waves as they propagate through the structure, and RF reflections from mismatched impedances in the electrode.

Determining the value of the lumped-element cut-off length l_{LE} for a transmission line is not an exact science. One expression for this, which deals purely with the transmission-line losses and assumes the resistance, capacitance and inductance are all frequency independent, is given as [81]

$$l_{LE} = \left| \frac{\Delta}{\sqrt{(i\omega L + R) \cdot i\omega C}} \right|. \quad (2.23)$$

In this context, R is the transmission-line series resistance per metre, L is the transmission-line series inductance per metre, and C is the transmission-line shunt capacitance per metre, while Δ is an arbitrary constant that is typically set to $\frac{1}{4}$ or, more stringently, $\frac{1}{10}$. For our application, at the high-frequencies in question (>1 GHz), Equation 2.23 is heavily influenced by the value of the inductance term, which heralds the arrival of the LC, skin-effect and dielectric regimes in the transmission line. However, as L is not an immediately accessible quantity, making an accurate estimation of the lumped-element cut-off length is difficult with this equation. This is also complicated by the fact that R and C can vary with the doping levels and other parameters of the phase shifter structure.

Since the boundary of the lumped-element region is by definition a rather fuzzy concept, a simpler yet sufficient estimation of l_{LE} is obtained by requiring the device to be acceptably small relative to the wavelength of the RF signal passing through it [82, 64]. A suitable expression for this is

$$l_{LE} = \left(\frac{\lambda_{RF0}}{n_{RF}} \right) \cdot \Delta, \quad (2.24)$$

where λ_{RF0} is the free-space RF signal wavelength (given by the ratio of the speed of light in vacuum to the signal frequency), n_{RF} is the RF refractive index or group index of the electrical transmission line, and the arbitrary constant Δ determines how stringent the restriction on the length is to be. Δ represents the acceptable ratio of the transmission line's length to the wavelength of the RF signal⁷. In this context, it can also be interpreted as the fraction of a cycle that the propagating RF signal completes before reaching the end of an electrode of length l_{LE} . For example, in the case of $\Delta = 1/4$, the signal at the start of the electrode is a quarter cycle or 90° out of phase with the signal at the end of the electrode at any given time.

Figure 2.13 plots Equation 2.24 for the two values of Δ mentioned above, assuming $n_{RF} \approx 4$. Clearly, the choice of Δ has an impact on the lumped-element cut-off length, with a 1.5 mm MZM remaining lumped until 5 GHz or 12.5 GHz, depending whether Δ is set to $\frac{1}{4}$ or $\frac{1}{10}$. The exact cut-off frequency for a given length is made less clear by the fact that MZMs designed for operation as lumped devices often have their RF contact points placed in the middle of the electrodes rather than at the end, to minimise the distance travelled by the RF signal in a

⁷Equation 2.23 can be seen to be related to this method if the transmission-line propagation coefficient γ is recognised in its two forms: $\gamma(\omega) = \alpha + i(2\pi/\lambda)$ and $\gamma(\omega) = [(i\omega L + R)(i\omega C)]^{1/2}$.

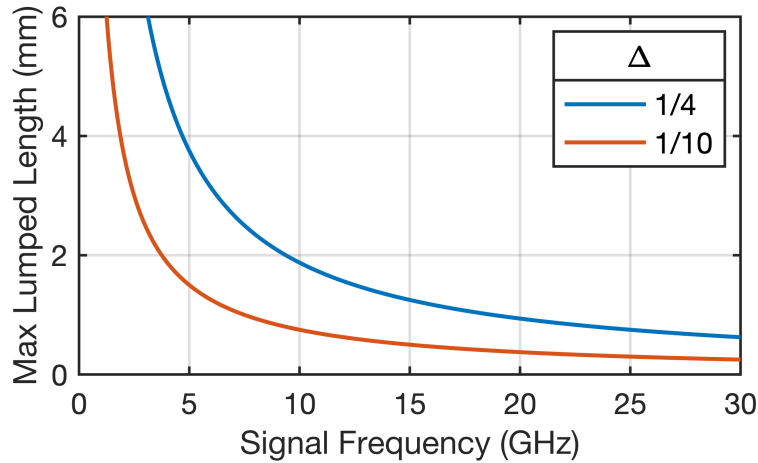


Figure 2.13: Lumped-element cut-off length plotted against signal frequency for two values of Δ .

given direction. In this case, the electrical length is considered to be “halved” for the purposes of the cut-off analysis [83, 84]. For a 1.5 mm MZM with a middle contact point, this brings the electrical length to 0.75 mm, corresponding to a lumped frequency cut-off of 10 GHz or 25 GHz for the two cases of Δ shown. Regardless of the finer details of the parameters used, it should be recognised that the boundary for the lumped-element regime is not a hard “cut off” at all. Rather, the bandwidth will be gradually penalised as the device is lengthened.

Shortening an electrode to operate as a lumped element does not stop the RF signals propagating along the line, receiving some attenuation as they go, and reflecting back from any impedance mismatch at the end load. Both a forward and backward mode of propagation exist, but the short length results in the temporal superposition of the forward and backward waves, creating the illusion of a instantaneous connection between the RF source and the load. In other words, *a transmission line is always a transmission line*. It is worth keeping in mind a quote in this regard from the excellent book by Johnson and Graham [81]:

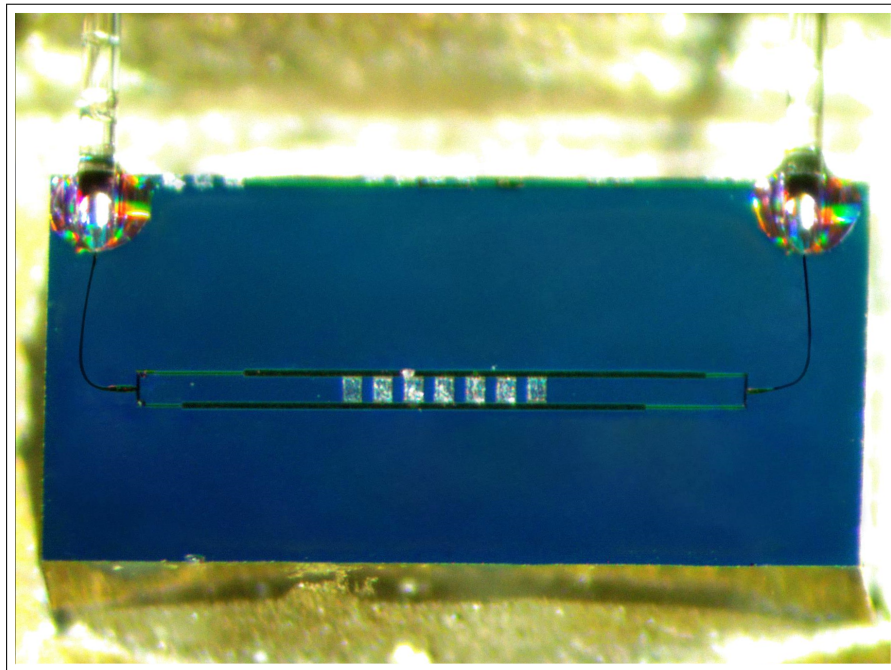
The classification of a transmission line in the lumped-element region does not determine how the line is going to act. It determines merely how the line may be analysed.

By extension, the classification also gives a rule of thumb for the design of the electrodes to be used in a modulator. MZMs that are considered to operate below l_{LE} are typically designed with simpler, lumped-style electrodes, which are viewed as a simple capacitor. These are referred to as lumped MZMs. In

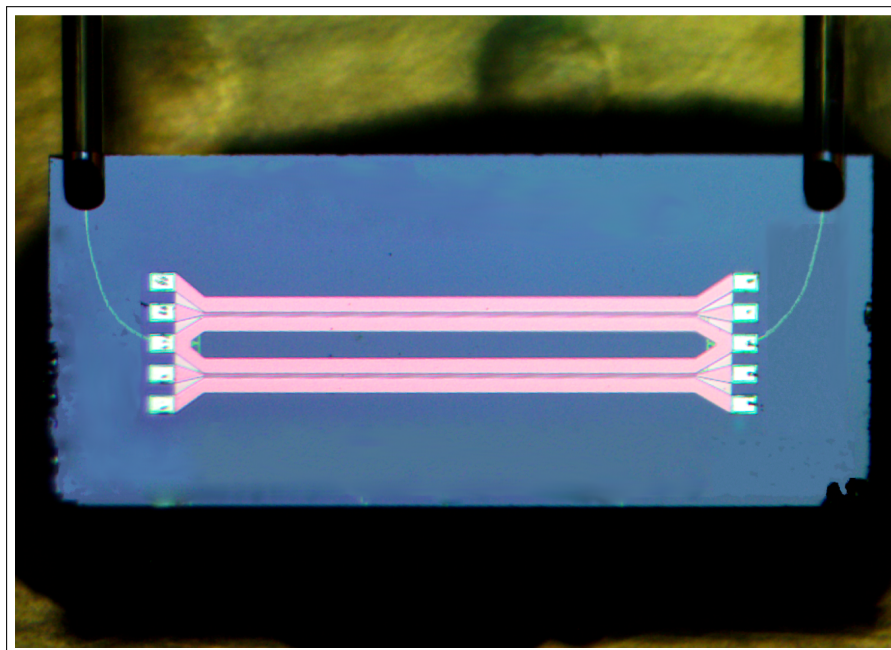
contrast, MZMs that are determined to operate beyond l_{LE} are designed with careful optimisation of the transmission line parameters; in particular, the RF attenuation coefficient, the RF group index, and the characteristic impedance of the electrode. These devices are referred to as travelling-wave MZMs (TW-MZMs).

Figures 2.14(a) and 2.14(b) show micrographs of a lumped SiP MZM and TW-MZM, respectively, that were used during the course of experiments in this thesis. Both devices are designed as dual-drive MZMs, and were fabricated as part of IMEC's ISIPP25G multi-project wafer process. These devices are semi-packaged, with planar optical coupling of waveguides completed by the packaging group at Tyndall National Institute. For clarity, extraneous bondpads and unused separate waveguides are mostly edited out of the images. More images are available in Appendix D. The lumped MZM has noticeably less wide electrodes, which are formed by deposition of thin ohmic metal contacts along the top of the waveguide. This provides a significant benefit in terms of device footprint on the chip. To deliver the RF signal to the device, small metal bondpads are placed nearby and connected electrically to the middle of the appropriate electrodes. These are necessary as the electrodes are usually only a few micrometres in width, and are otherwise difficult to contact with probes, bond wires, or flip-chip solder balls. Five of the central pads in Figure 2.14(a) are used for testing this lumped MZM, with the pads connected to the ground (G) and signal (S) electrodes in a GSGSG format. The sixth and seventh pads (located on the far left and right of the row) are redundant. In contrast to the narrow electrodes of the lumped MZM, the electrodes of the TW-MZM in Figure 2.14(b) are fabricated as much wider, and slightly thicker, co-planar waveguides. These are designed to support the propagation of high-speed signals in the transmission line. The RF signal is applied to the GSGSG bondpads at one end of the electrode and co-propagates with the optical signal, before being terminated at the far-end bondpads with a resistive load that ideally matches the characteristic impedance of the electrodes.

The active length of each MZM is 1.5 mm. Whether a lumped MZM of this length can be considered to be within the lumped-element regime depends on the value of Δ chosen for use in Equation 2.24. The true test of such an MZM's capabilities, therefore, is to evaluate it in system-level experiments. This analysis is done using multiple driver schemes as part of this thesis.



(a)



(b)

Figure 2.14: Micrographs of (a) dual-drive SiP Lumped MZM, and (b) dual-drive SiP TW-MZM, taken at a similar scale.

2.5 Power Consumption and Drive Voltage

The power consumption of TW-MZMs and lumped MZMs differs significantly due to the presence of (or lack thereof) a termination resistor. A discussion of their energy demands therefore requires a proper description of their equivalent circuits in order to identify where voltage drops and current is flowing. Such a treatment will be given in Section 4.5, but a simplified analysis is presented here to give an initial estimate of the power consumption and driver requirements as context for using lumped MZMs over TW-MZMs.

In this section, the voltages discussed are in terms of root-mean-square (RMS) values rather peak-to-peak measurements, unless otherwise stated. For square waves, this corresponds to the amplitude of the wave. The power consumption discussed is the dynamic power consumption of the modulator, which arises only from the applied RF voltage. The bias voltage is assumed to cause no additional energy consumption in either the lumped or travelling-wave cases. For lumped MZMs, this holds as the capacitive nature of their operation means that no significant static current is present. Although the bias voltage does normally drop over the termination resistor in TW-MZMs, a DC block or capacitor can be inserted between the transmission-line electrodes and the termination in order to prevent any static power consumption. It is simple to add a discrete DC block component in a lab setting if termination is obtained via an RF probe with an attached RF terminator [85, 86]. However, for a fully-integrated device, suitable capacitance must be added via on-chip elements which requires significant chip area [87].

Figure 2.15(a) shows a typical transmission line description of the drive circuitry

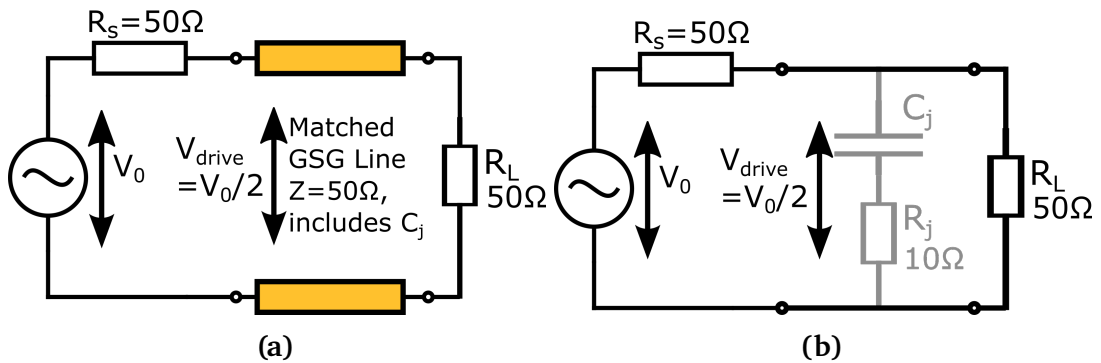


Figure 2.15: (a) Transmission line diagram of a TW-MZM with 50Ω driver and matching termination resistor. (b) Simple equivalent circuit diagram for the same structure.

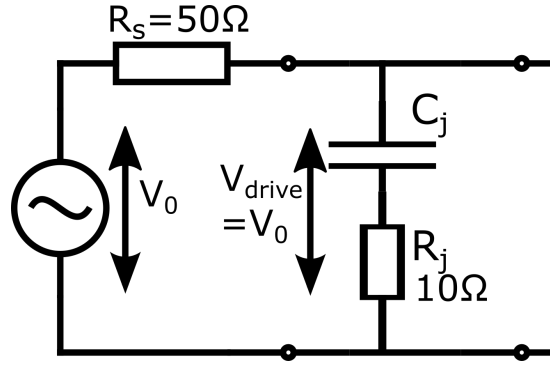


Figure 2.16: Equivalent circuit of a lumped MZM with 50 Ω driver.

for one arm of a TW-MZM [88], including a 50 Ω source impedance R_S and a 50 Ω termination resistor R_L placed as a load across the end of the MZM electrodes. V_0 is the amplitude of the open-circuit voltage of the voltage source; the voltage falling across the MZM arm will be split according to the ratios of R_S and R_L . Assuming R_S , R_L and the characteristic impedance of the transmission line are all matched, the voltage V_{drive} across the MZM arm will be $V_0/2$. For clarity, Figure 2.15(b) shows an equivalent circuit model of the same arm, representing the SiP phase shifter by its junction capacitance capacitor C_j (typically in the order of hundreds of fF) and junction resistance R_j (typically around 10 Ω). In terminated TW-MZMs, the power consumed in an arm of the structure is normally assumed to be dominated by the resistive termination at the end of the line rather than the charge-discharge cycle of the PN junction phase shifters. Hence, the PN junction is often ignored for the purposes of power consumption estimates, and the transmission line and load impedance are replaced with a single resistor of 50 Ω in the equivalent circuit [89]. Then, following Ohm's law, and if R_S and R_L are equal, the total power consumption in the TW-MZM will be

$$P = \left[\frac{V_{drive}^2}{R_S} + \frac{V_{drive}^2}{R_L} \right] = \frac{2V_{drive}^2}{R_L}, \quad (2.25)$$

where the drive voltage is assumed to be a rectangular NRZ signal with equiprobable ones and zeros, and a 50 % duty cycle.

Figure 2.16 shows an equivalent circuit for one arm of a lumped MZM with a typical 50 Ω driver. The phase shifter is represented by a capacitance C_{arm} lying in series with its resistance R_{arm} . Typically, as an approximation, the energy consumption of a lumped MZM is assumed to follow that of a capacitor being fully charged and discharged by a DC source. In this case, the energy dissipated in charging and discharging such an RC series circuit does not depend on the

size of the resistor, and so follows the usual expression for energy stored in a capacitor. As energy is only consumed during bit transitions (when the capacitor is charging and discharging), the power consumption depends on the symbol rate f_{sym} of the RF signal driving the MZM. For an NRZ signal, the MZM's power consumption will then be given by

$$\begin{aligned} P &= \frac{1}{2} \cdot \left(\frac{1}{2} C_{arm} V_{drive}^2 \right) \cdot f_{sym} \\ &= \frac{1}{4} C_{arm} V_{drive}^2 \cdot f_{sym}, \end{aligned} \quad (2.26)$$

where the first factor of $\frac{1}{2}$ accounts for the fact that symbol transitions occur 50 % of the time. Crucially, with the absence of a termination resistor, the drive voltage across the phase shifter V_{drive} is now equal to the source voltage V_0 , rather than being halved. This allows lower voltages to be set at the driver to achieve the same phase shift as a TW-MZM.

Finally, to see the energy required per bit for a given drive format, the respective power consumptions of Equations 2.25 and 2.26 should be divided by the symbol rate f_{sym} . These are shown below for the two drive cases:

$$\text{Travelling-wave MZM:} \quad E_{perBit} = \frac{1}{f_{sym}} \cdot \frac{2V_{drive}^2}{R_L}, \quad (2.27)$$

$$\text{Lumped MZM:} \quad E_{perBit} = \frac{1}{4} C_{arm} V_{drive}^2, \quad (2.28)$$

To make a fair comparison of the energy consumption of the two cases, it must be ensured that each drive setup produces the phase shift in the MZM arm, i.e., that V_{drive} is the same in all cases. This means that V_0 will need to be doubled for the TW-MZM compared to the lumped MZM. This comparison is shown in Figure 2.17(a) for the two cases, where the energy per bit is plotted against the symbol rate for an NRZ signal. The capacitance of the silicon phase shifter is assumed to be 500 fF and the peak-to-peak drive voltage is kept constant at 1 V for each case. At low symbol rates, the travelling wave design is seen to cost significantly more energy per bit than the lumped design, reaching well above 1 pJ/bit for symbol rates lower than 10 GHz while the lumped design stays fixed at approximately 30 fJ/bit. As the symbol rate increases, however, the TW-MZM's energy efficiency increases due to inverse relationship to f_{sym} as described by Equation 2.27. To compare this change properly, Figure 2.17(b) plots the ratio of the energy consumption of the TW-MZM to the lumped MZM. Despite the

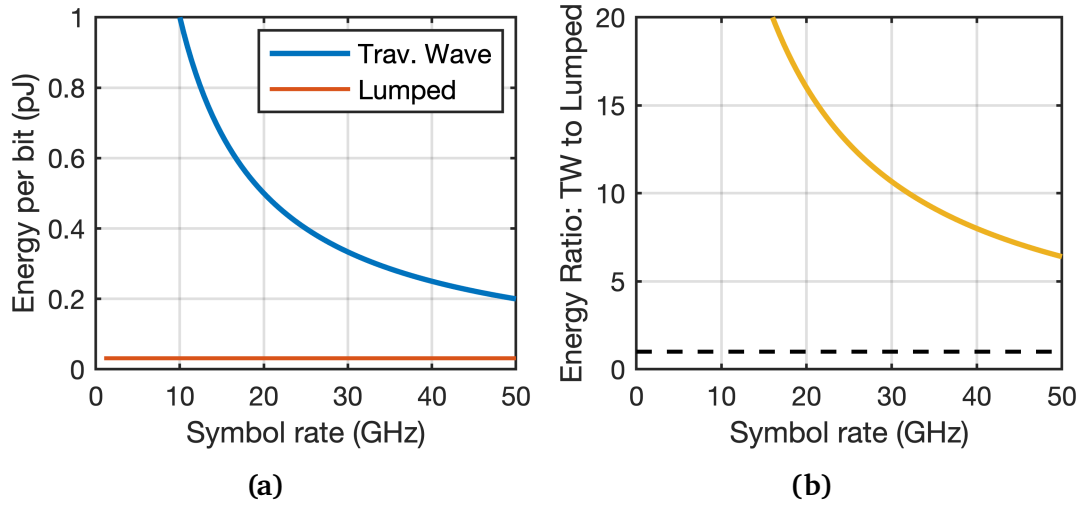


Figure 2.17: (a) Energy per bit for a single MZM arm for the two drive scenarios. (b) The ratio of the TW-MZM energy consumption against that of the lumped MZM.

improvement in the TW-MZM's energy per bit with symbol rate, the lumped MZM remains more than 10 times more energy efficient than the TW-MZM for symbol rates up to 30 GHz.

Of course, there is a large caveat here—this comparison does not give any consideration to the effects of bandwidth and circuit parasitics, and is essentially a quasi-DC treatment of the power consumption. The energy dissipation in a given equivalent circuit can look very different at high frequencies in comparison with low. Therefore, as mentioned at the start of this section, a more thorough analysis of the lumped MZM and TW-MZM's equivalent circuits and power consumption will be given in Section 4.5.

This analysis also does not take into account any overhead in power consumption that is introduced by the electrical driver circuitry. While a detailed investigation into any given driver's power consumption requires analogue design knowledge and simulations that are beyond the scope of this thesis, an overview of suitable driver types and their general energy trends is given in Section 4.6.4.

Finally, in reality, lumped MZMs are generally far more restricted in bandwidth than TW-MZMs and so cannot easily reach symbol rates in the tens of GHz. Careful design of lumped MZMs and their drivers is therefore required to maximise bandwidth, as will be discussed in Chapter 3. Equalisation is a possible method of increasing the bandwidth of lumped MZMs, although equalisation circuits do bring an additional amount of energy consumption to

communication links. The possibility of using equalisation for lumped MZM designs is analysed through simulations and experiments in Chapter 4.

2.6 State of the Art

The earliest optical modulators based on silicon waveguides emerged in the 1980s [33, 90]. These devices were based on carrier injection and, hence, were very slow, with data rates in the order of megahertz. Carrier injection designs typically have low bandwidths due to long minority carrier lifetimes in silicon [79]. Thus, it was a carrier accumulation design by Intel, rather than an injection design, that first broke the 1 GHz bandwidth barrier [91]. Such accumulation designs are not limited by carrier lifetimes, but instead dominated by the resistance and capacitance of the device. This milestone was quickly followed by another, with an improved performance of 10 Gb/s NRZ demonstrated by the same group the following year [92]. The move to truly high-speed silicon photonic MZMs was due to a carrier depletion design proposed in 2005 [93], with device modelling predicting groundbreaking bandwidths of over 50 GHz (albeit in the absence of any bandwidth limitations that would be inevitably introduced by electrical drivers). In 2007, Intel demonstrated a depletion-mode SiP TW-MZM that operated at 30 Gb/s NRZ [94], and, later, 40 Gb/s NRZ [95]. Since then, numerous carrier-depletion TW-MZMs operating in the region of 40 Gb/s to 60 Gb/s NRZ data rates have been demonstrated [96, 97, 98, 99, 86], with some reaching as high as 90 Gb/s NRZ [100]. Typically, these devices are reported as being driven with high peak-to-peak RF voltages of 4 V to 7 V, achieving extinction ratios in the region of 3 dB to 6 dB. One reason for the large required drive voltage is their operation as terminated TW-MZMs, which halves the modulation effect as described in Section 2.4.

Lumped MZMs offer more efficient use of drive voltage when compared with TW-MZMs, but typically offer lower modulation bandwidths due to their lack of impedance-matching termination resistors. Typically, silicon photonic lumped MZMs have been reported as having very low bandwidths unless their lengths are kept very short [101], which in turn increases their V_π and required drive voltage. Because of this, lumped MZMs have been relatively out-of-favour in the silicon photonics literature throughout the 2010s in comparison with TW-MZMs. Recently, however, research interest in lumped SiP MZMs has renewed due to their potential for low-power consumption. Short-length, lumped SiP MZMs driven with low voltages have shown promise for use in low-extinction-

ratio, chip-to-chip interconnects for high-performance computing applications [102]. Furthermore, effort has been made to optimise lumped electrode designs to maintain bandwidth over longer-lengths of MZMs [103, 104, 105]. Some interest has also been shown in adding equalisation in the driver circuits to boost lumped MZM bandwidths [106], which is also explored in Chapter 4 of this thesis.

2.7 Summary

This chapter gives an overview of Mach-Zehnder modulators within the context of silicon photonics, in order to establish some fundamentals needed for the rest of the thesis. Section 2.2 provides the key equations that govern the modulation principle of MZMs, starting from the required Δn_{eff} for a given phase shift, and moving through to the drive formats that can be used for the devices. Ways of tuning the operation point on the MZM's transfer function through thermal, electrical or wavelength means are identified; wavelength tuning is noted as a convenient method which is used throughout the experiments later in this thesis. Section 2.3 describes the various methods of achieving phase shift and intensity modulation in silicon photonic MZMs, with the free-carrier plasma-dispersion effect in reverse-biased PN-junctions established as the effect of choice for this work, as in many others. Section 2.4 discusses some key differences between lumped and TW-MZMs, offsetting the advantages of lumped MZMs' smaller physical footprint and drive voltage against their bandwidth limitations. The need to analyse the performance of relatively long-length lumped MZMs through detailed experiments and simulations is emphasised. Section 2.5 presents the typical expressions that are used to estimate the power consumption of lumped MZMs (which is typically assumed to be purely capacitive in nature) and TW-MZMs (which is viewed as being purely resistive and therefore much larger). The difference in their respective power consumptions is the justification for why lumped MZMs were investigated further in this thesis; a more detailed comparison of the two MZM designs will be given later in Chapter 4. Finally, in Section 2.6, the trends and state of the art in silicon photonic MZMs are discussed.

Chapter 3

Optimisation of Lumped SiP MZMs

3.1 Introduction

This chapter describes the characterisation, modelling, and optimisation of lumped SiP MZMs. A method of measuring the phase shift and absorption change against voltage for silicon phase shifters is given for unbalanced MZMs. A description is given of how to characterise the frequency response of a system using the equipment that would be typically present in an optical testbed, without needing a vector network analyser. This method is particularly useful given the investigations on predistortion that will be carried out later in Chapter 4. An accessible model for the phase shift and absorption of SiP depletion-mode phase shifters is described, which involves co-simulation of the electrical and optical properties of the PN junctions. This model is used to investigate the effect of various doping levels in the PN junctions on V_π , insertion loss, and other system parameters. The effect of the doping profile on the RC bandwidth of lumped MZMs is modelled, and the resulting impact of driver impedance is shown alongside this. The benefits of a low-impedance, switched capacitor drive scheme on the bandwidth of lumped SiP MZMs is discussed, and the phase shifter model is used to optimise the bias and doping levels of the PN junctions for PAM4 and PAM8 modulation. Finally, a full system-level demonstration of a SiP link consisting of a lumped, switched-capacitor SiP MZM and an SiP photoreceiver is shown to operate successfully with low-power 30 Gb/s PAM4 transmission.

3.2 Characterisation of SiP MZMs

This section presents measurements and an overview of how key parameters of SiP MZMs (current, phase shift, absorption, and bandwidth) vary with voltage. These measurements are particularly relevant for SiP MZMs, given that the LiNbO₃ phase shifters in traditional MZM designs typically have a linear phase-voltage relationship, and have no absorption or bandwidth changes with voltage.

3.2.1 Voltage-Current Characteristics

The first characterisation that should be made with a silicon photonic MZM is its VI curve, i.e., the relationship between the DC voltage dropped across the PN junction phase shifters and the resulting current draw. This is done to check that the phase shifters are electrically functional, and also to ensure that the voltage is being applied in the correct direction for reverse bias in subsequent experiments.

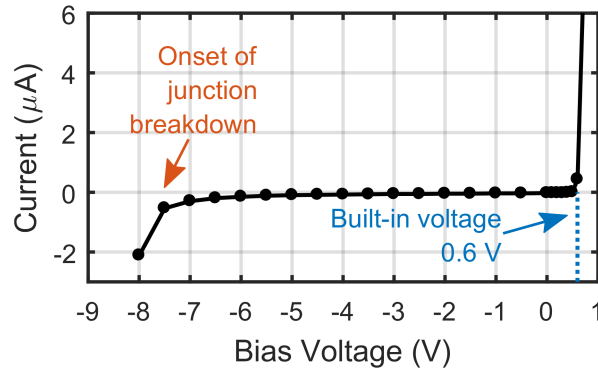


Figure 3.1: VI curve for one arm of the lumped MZM, showing diode behaviour.

Electrical contact is made with each arm of the MZM in turn using suitable probes (GSG-style RF probes were used in this case in preparation for later high-speed experiments, but DC probes would suffice for this characterisation). A Keithley source meter is used to accurately provide and measure voltage and current. The voltage is swept carefully in either direction around zero, with the resulting current plotted at each step until the direction of forward and negative bias is determined for the diode, as shown for one arm in Figure 3.1. Forward bias in a silicon PN junction is indicated by a sharp increase in current at a voltage in the region of 0.5 V to 1 V known as the built-in voltage (V_b). In this case, this is seen clearly at 0.6 V. The PN junction is then under reverse bias for any voltage lower than this. Under reverse bias, until approximately -7 V, the current drawn is very small (<100 nA). Beyond this point, the current starts to increase rapidly,

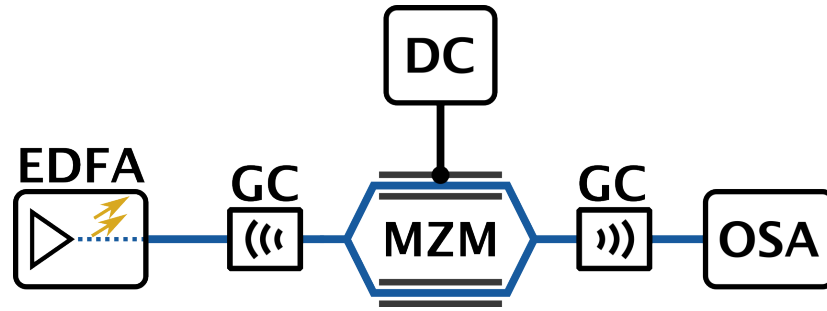


Figure 3.2: Setup for measuring phase shift and absorption characteristics of SiP MZM against voltage.

indicating the onset of breakdown in the junction due to avalanche or Zener breakdown mechanisms [107]. As the rapidly increasing current beyond this point can cause irreversible thermal damage (unless limited by a suitable external resistance), this voltage region is generally avoided in silicon modulators.

3.2.2 Measuring Phase Shift and Absorption with an OSA

The measurement of the voltage-induced phase shift and absorption change from the phase shifters in any MZM requires the MZM to be unbalanced as described in Section 2.2.6, with a fixed path length difference in the two arms. Figure 3.2 shows the measurement setup for characterising the phase shift and absorption characteristics of a depletion mode phase shifter in a 1.5 mm SiP MZM by measuring its transmission spectrum on an OSA. Amplified spontaneous emission (ASE) from an erbium-doped fibre amplifier (EDFA) is used as a broadband light source, and is passed into the SiP MZM using an optical fibre aligned to the on-chip grating coupler (GC). As the EDFA outputs unpolarised light, no polarisation controller (PC) is needed before the input GC of the MZM.

A DC voltage is applied to one arm of the MZM, and is swept for values of reverse bias¹ from 0.6 V to -7 V. The output of the MZM is passed into an optical spectrum analyser (OSA) using another optical fibre. The wavelength dependence of the EDFA's ASE spectrum is calibrated out of the measurements, and the resulting MZM output spectrum is recorded for each value of voltage, as shown in Figure 3.3. As the MZM is unbalanced, as described in Section 2.2.6, the spectrum has a noticeable sinusoidal variation with wavelength. For legibility, the graph shows only one half cycle of the spectrum and has been normalised to

¹A clarification on wording: the reverse bias is “increased” as it moves towards greater negative voltages. Slight positive values of voltage are still considered as reverse bias, provided they are below the built-in voltage (see Section 3.3.1).

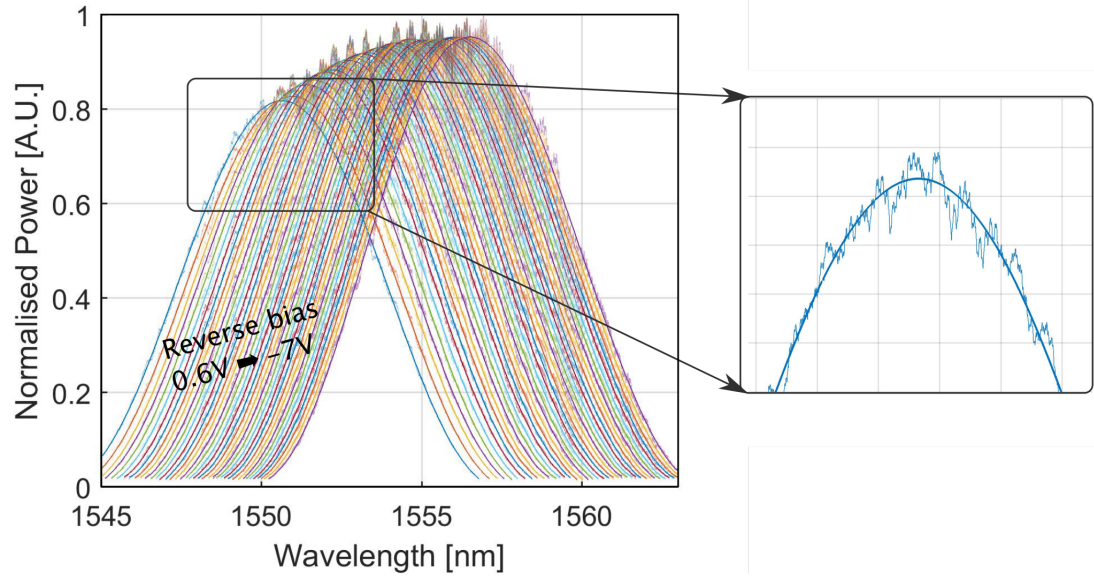


Figure 3.3: Measured transmission spectrum dependence on voltage for the SiP MZM. The zoomed section shows a single peak to highlight the Fabry-Perot resonance and the need for a fitted curve.

the peak recorded power. Increasing the reverse bias voltage causes the spectrum to shift noticeably to the right, and also causes the peak transmitted power to rise.

Unfortunately, there are noticeable periodic ripples or spurious peaks in the spectrum which distort the data. These are characteristic of Fabry-Perot resonances formed by internal reflections of light within the MZM from imperfect device structures such as the grating couplers or splitters. To determine the culprit for this effect, the length l of the unwanted Fabry-Perot resonator is determined from $FSR_{fp} = \lambda^2 / 2n_g l$ [39], where FSR_{fp} is the free-spectral range of the Fabry-Perot resonator as measured from the period of the ripples (0.5 nm), λ is the optical wavelength (1550 nm), and n_g is the group index of waveguide (value of 3.84). This equation is essentially Equation 2.8 with a factor of 2 in the denominator to account for the round-trip in the path. The calculated length of the cavity is approximately 0.6 mm, which corresponds to the distance between the grating couplers and waveguide splitters on the MZM. In spite of their noticeable effect on the spectrum, these peaks do not affect the performance of the MZM when used in the transmission links in this thesis, as the optical wavelength is kept fixed.

To identify the true peaks and FSR of the MZM spectra, polynomial fits are applied to the data as shown in Figure 3.3, and the peak of each fitted curve

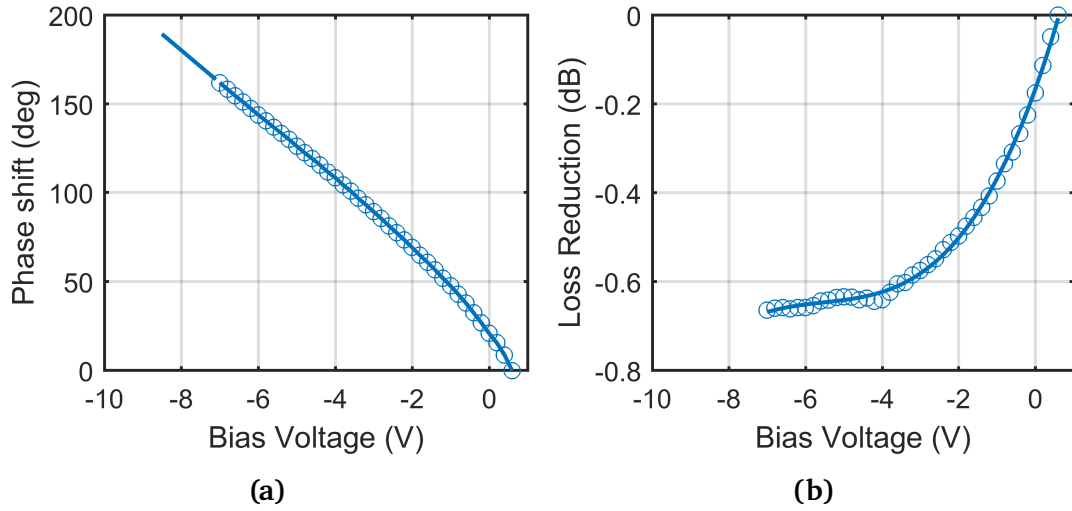


Figure 3.4: (a) Measured phase shift and (b) change in absorption against bias voltage, for a 1.5 mm SiP depletion-mode phase shifter.

is extracted. The phase shift produced by the silicon phase shifter can then be calculated from

$$\Delta\phi = \frac{\pi\Delta\lambda_{pk}}{FSR}, \quad (3.1)$$

where $\Delta\lambda_{pk}$ is the shift in peak wavelength due to the applied voltage [94, 108]. The FSR is measured between two successive peaks of the spectrum at a fixed bias as 13.25 nm, which, using Equation 2.8, corresponds to a length imbalance of 73 nm. The resulting phase shift against voltage is plotted in Figure 3.4(a). The phase shift is noticeably non-linear with voltage, especially at small reverse biases. This is to be expected for a depletion-mode phase shifter, as the width of the depletion region varies with the square root of the applied voltage. Extrapolating the measured data shows that a full π phase shift (180°) is achieved at a bias of -8.5 V. The change in the phase shifter's absorption is measured from the variation in the transmission spectrum's peak power with voltage.

Figure 3.4(b) plots the absorption change in terms of dB. The absolute values for the free-carrier absorption could be determined if the values for the other sources of loss in the MZM, such as the coupling loss and scattering loss from sidewall roughness, were tightly known. This can be achieved by coupling to and from a passive waveguide of the same length in order to measure the loss in the absence of doping. Unfortunately, such calibration waveguides were not accessible on these SiP chips, so the absolute free-carrier absorption could not be determined (though simulated values in Section 3.4.2 place it in the region of 0.5 dB/mm to 2 dB/mm, depending on the doping levels used in the device).

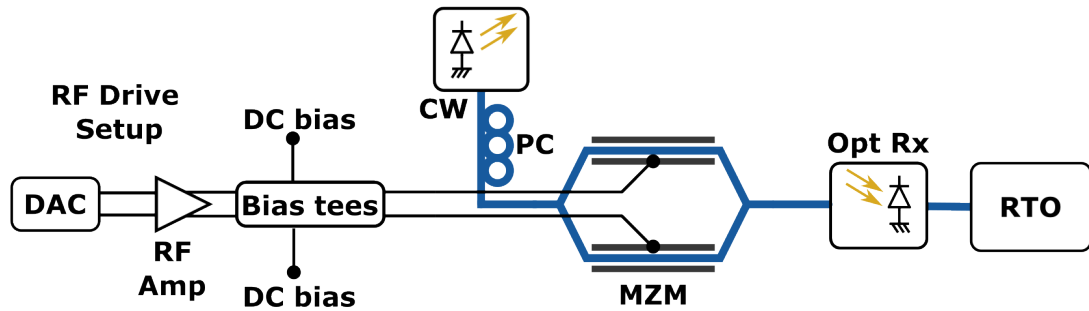


Figure 3.5: Typical optical testbed for systems tests of an MZM, suitable also for frequency response measurements using SPHS.

The absorption clearly varies non-linearly with voltage, and is seen to decrease for greater reverse biases. The overall variation in the absorption is quite small, however, with less than 0.7 dB change across the measured voltage range. This implies that it will have minimal effect on system performance, which is verified in Section 3.4.2.

3.2.3 Measuring Bandwidth with an AWG and Oscilloscope

The frequency response of a device or system is generally measured with a vector network analyser (VNA). They operate by stepping up the frequency of sinusoidal signal transmitted through a system, and measuring the amplitude and phase of the corresponding received sinusoid. High-spec VNAs can measure to 100 GHz or more, but are bulky, tedious to connect, and require disturbing the experimental setup. Therefore, it would be convenient if the frequency response of a device in a communication testbed could be measured instead by using an arbitrary waveform generator (AWG), DAC and real-time oscilloscope (RTO) that could be typically present, as shown in Figure 3.5. Generating a sine wave to mimic the output of a VNA is possible with an AWG, but loading the signal pattern to the memory of the device is too slow to allow a stepped-tone frequency response measurement of any significant resolution or bandwidth. A single pattern in the time domain that covers the frequency range of interest is thus required.

It is well known that the inverse Fourier transform of a flat power spectrum across all frequencies in the frequency domain is a delta function in the time domain. Of course, generating a true delta function is impossible with electronic hardware due to inevitable limits on rise/fall times, sample resolution and vertical resolution. Even when we consider that we do not need to cover all possible

frequencies to measure the response of our system, this remains a problem. A limited, rectangular distribution of frequencies covering our range of interest will require a sinc function in the time domain, rather than a delta function, but this is still a difficult pulse to implement. For example, to see how sample resolution alone could be an issue, assume that the maximum frequency to be measured is 50 GHz and that we need a sample frequency of 6 times this to accurately represent the sinc function. We would then require a 300 GS/s DAC, which is beyond reasonable hardware capabilities.

However, a flat spectrum in the frequency domain can be achieved with an alternative to the delta function. In 1970, in the course of a paper discussing how to specify the phase angles of individual harmonics to achieve low-peak-factor periodic signals, Schroeder [109] noted that the resulting signals resembled “a chirped FM signal with monotonically increasing instantaneous frequency”. We will call these “Schroeder-phased harmonic signals” (SPHS). He then derived an expression for the phase angles that would cause these signals to have flat power spectra. This derivation was corrected slightly in 1978 by Flower, who applied flat-spectrum SPHS to measure the frequency response of systems [110, 111]. The requirements on the signal for this condition are shown next.

Consider a periodic, wide-band signal $r(t)$, with period T . It can be expressed in terms of its Fourier series as

$$r(t) = \sum_{k=1}^N (2p_k)^{\frac{1}{2}} \cos\left(\frac{2\pi kt}{T} + \theta_k\right), \quad (3.2)$$

where N is the total number of harmonics, p_k is the relative power of the k th harmonic, i.e., $\sum_{k=1}^N p_k = 1$, and θ_k is the phase of this harmonic. To obtain an SPHS (a low-peak-factor signal), the phases of the harmonics must follow

$$\theta_k = 2\pi \sum_{i=1}^k ip_i, \quad k = 1, 2, \dots, N. \quad (3.3)$$

In a flat spectrum, each harmonic has equal power content, i.e., $p_k = 1/N$ for all k . This gives the phase requirement in Equation 3.2 for $r(t)$ to be a flat-spectrum SPHS as

$$\theta_k = \frac{2\pi}{N} \sum_{i=1}^k i = \frac{\pi(k^2 + k)}{N}. \quad (3.4)$$

The linear k term in equation Equation 3.4 acts only as a delay in the signal, and is sometimes omitted in the expression. With the above stipulation for the phases

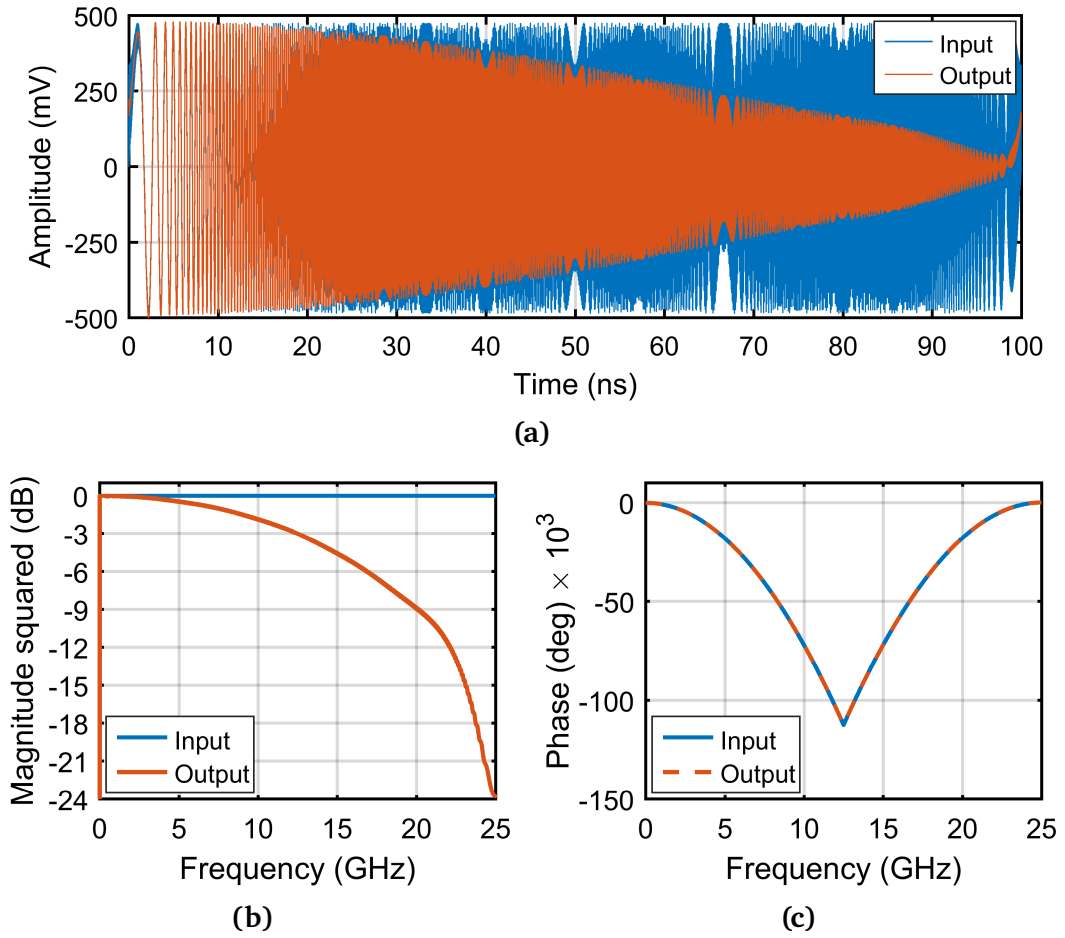


Figure 3.6: Simulated input and output for flat-spectrum SPHS passed through an ideal fourth-order, 12.5 GHz Bessel filter, shown as: (a) time domain waveform, (b) amplitude-frequency response, and (c) phase-frequency response.

of the harmonics, it is possible to conveniently generate flat power spectra for frequency response measurements up to the Nyquist frequency of the DAC. Of course, the AWG, DAC and oscilloscope will have their own frequency roll-offs that play a part in the system, but it is easy to digitally calibrate these out in the course of the measurements.

To demonstrate how SPHS are used in frequency response measurements, Figure 3.6(a) plots, in the time domain, one period of a simulated, flat-spectrum SPHS before and after transmission through a channel with an ideal fourth-order, 12.5 GHz Bessel response. This particular SPHS is designed for a frequency response measurement from 10 MHz to 25 GHz in steps of 10 MHz, with a sample frequency of 50 GHz, and has a peak-to-peak voltage of 1 V. As expected, the input waveform has a low peak factor, and is seen to increase in instantaneous frequency over the course of its period. The output waveform after

the channel decreases in amplitude as the instantaneous frequency increases. After suitable application of a fast-Fourier transform, the amplitude-frequency and phase-frequency spectra can be obtained for each waveform, as shown respectively in Figures 3.6(b) and 3.6(c). The amplitude-frequency response, initially flat for the input, takes the expected form of a 12.5 GHz Bessel filter at the output. The phase-frequency plot is identical for the input and outputs in this case, as a Bessel filter by nature does not disturb the phase of a signal. The phase-frequency plot exhibits a minimum point at half the maximum frequency used, which is characteristic of SPHS inputs.

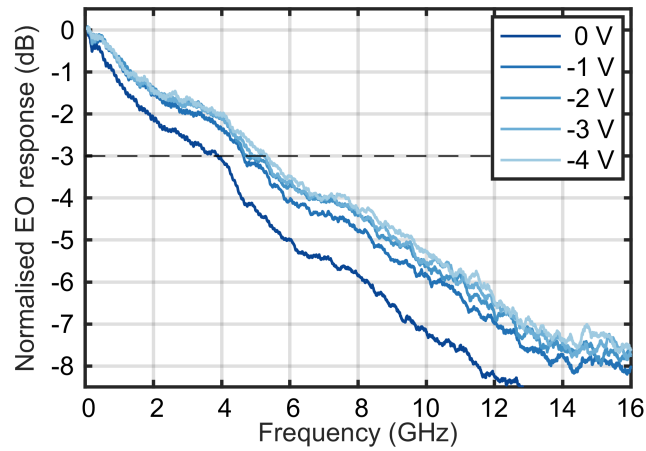


Figure 3.7: Frequency response of an unterminated SiP MZM at various bias voltages, measured using SPHS.

Using this approach in an experimental setup similar to Figure 3.5, Figure 3.7 shows the measured electro-optic frequency response of an SiP MZM at various bias voltages. This particular device is a 1.5 mm TW-MZM with no termination resistor applied at the end of the electrodes, which is used as part of the experimental work in Chapter 4. The peak-to-peak voltage of the SPHS used to drive each MZM arm in the measurement is 1 V. With reference to Figure 3.4(a), this voltage results in approximately 30° phase shift in each arm (less than 60° total). This is sufficiently small to operate in the linear part of the MZM's transfer function, and thus can be considered to be in the small-signal regime. The voltage in each arm is kept small enough to remain below the built-in voltage of the PN junction when biased at 0 V and so the device remains in depletion mode. The measured frequency response is bias dependent, with the 3 dB electro-optic bandwidth increasing from 4 GHz at 0 V to 5.1 GHz at -4 V. The measured bandwidth is low due to operating the device in a $50\ \Omega$ environment without termination; methods of boosting the bandwidth of unterminated/lumped MZMs using low-impedance drive schemes and equalisation will be discussed in

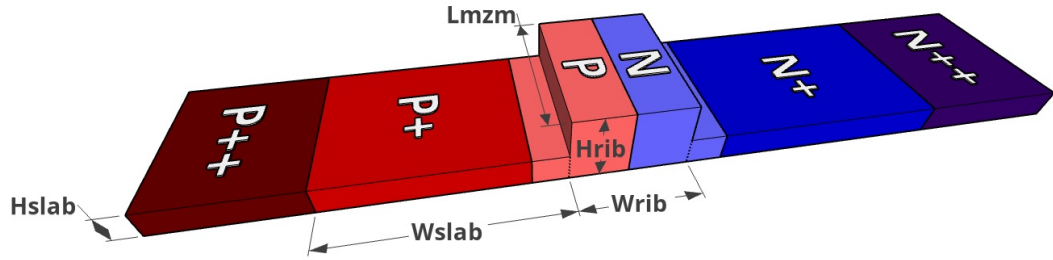


Figure 3.8: 3D diagram of a silicon p-n junction phase shifter section (not to scale), showing heavily doped ohmic contact regions (P/N++), intermediate doping regions for resistance reduction (P/N+) and the central doping region (P/N) in which the light propagates and refractive index change occurs. The depletion region between the P and N sections is not shown.

Section 3.4.1 and Chapter 4, respectively.

3.3 The Roll of Doping Levels in SiP MZMs

As mentioned in Section 2.3.2, silicon waveguides are doped to form the p-n junction phase shifters used in depletion-mode MZMs. The doping levels must be carefully chosen to balance the modulation efficiency, optical absorption and modulation bandwidth of the device. This section gives an overview of how this balance is modelled.

Figure 3.8 shows the doping profile of a typical lateral-junction depletion mode phase shifter. The lightest doping tends to be present in the central waveguide through which the light propagates, as heavy doping will cause too much optical absorption. At the edge of the rib waveguide, far from the optical field, very heavy doping is used to facilitate good ohmic contact between the p-n junction and the metal vias. The connecting sections of the slab can be left with the same doping as the central waveguide region, but ideally should be given a moderately heavy intermediate doping in order to reduce the series resistance of the phase shifter and thus improve the RC bandwidth, as will be shown in Section 3.3.3.

3.3.1 Modelling the phase shift and absorption of SiP MZMs

The process of determining the phase shift and and absorption produced by a silicon depletion-mode phase shifter requires co-simulation of the PN junction's optical and electrical properties. The optical simulation is needed to determine the distribution of the optical mode within the rib waveguide. A commercial

mode solver (*FemSIM* from Synopsys [48]) was used to simulate this, as discussed in Section 2.2.2. The mesh size used for the simulation was swept with increasing fineness until convergence in the output was obtained. The mode-solving approach used here gives strictly a two-dimensional description of the optical field, which is sufficient as the phase shifters can be viewed as homogeneous along their length. The use of FEM to solve the optical mode is relatively low in computational effort, particularly as it is only required to be done once if the waveguide cross-sectional geometry is fixed.

The electrical simulation is used to determine the distribution of carriers in the PN junction at specified voltage biases, which requires solving the Poisson equation, carrier continuity equations, and drift diffusion equations for each one of those voltages [78], while taking the junction geometry into account. These equations are solved through numerical techniques, such as FEM [112, 113, 114], which can be time consuming and tricky to implement in this context. Commercial technology computer-aided-design (TCAD) software packages are also available to do such calculations [115, 116, 117], but these can often be quite expensive to obtain. In addition, this depth of simulation is typically memory intensive and requires long computation times, as it must be recalculated for each voltage in the range of interest. This can slow down analysis and limit optimisation, given that the entire process must be repeated if the doping parameters are wished to be changed. Therefore, it is beneficial to reduce the simulation complexity by making it analytic, where possible, through appropriate simplification and assumptions, particularly when trying to identify trends and general characteristics of the silicon phase shift technology.

With this in mind, the approach in this thesis follows the treatment of [118], in which the optical mode is simulated rigorously, while the electrical simulation simplifies the free carrier distribution to a one-dimensional treatment through a simple analytic expression for the PN junction's depletion region. An overlap integral is then taken of the carrier distribution and optical mode for various values of applied voltage to determine the effective refractive index change and free-carrier absorption coefficient. A final integration along the length of phase shifter then provides the phase shift and optical loss. (A more thorough description of the model follows below.) This method is taken as a balance between the more simple treatment of [119], in which both the optical and electrical distributions are reduced to analytic expressions in one dimension, and the more recent and more involved treatment of [120], which expands the electrical description to a two-dimensional analytic model including fringing

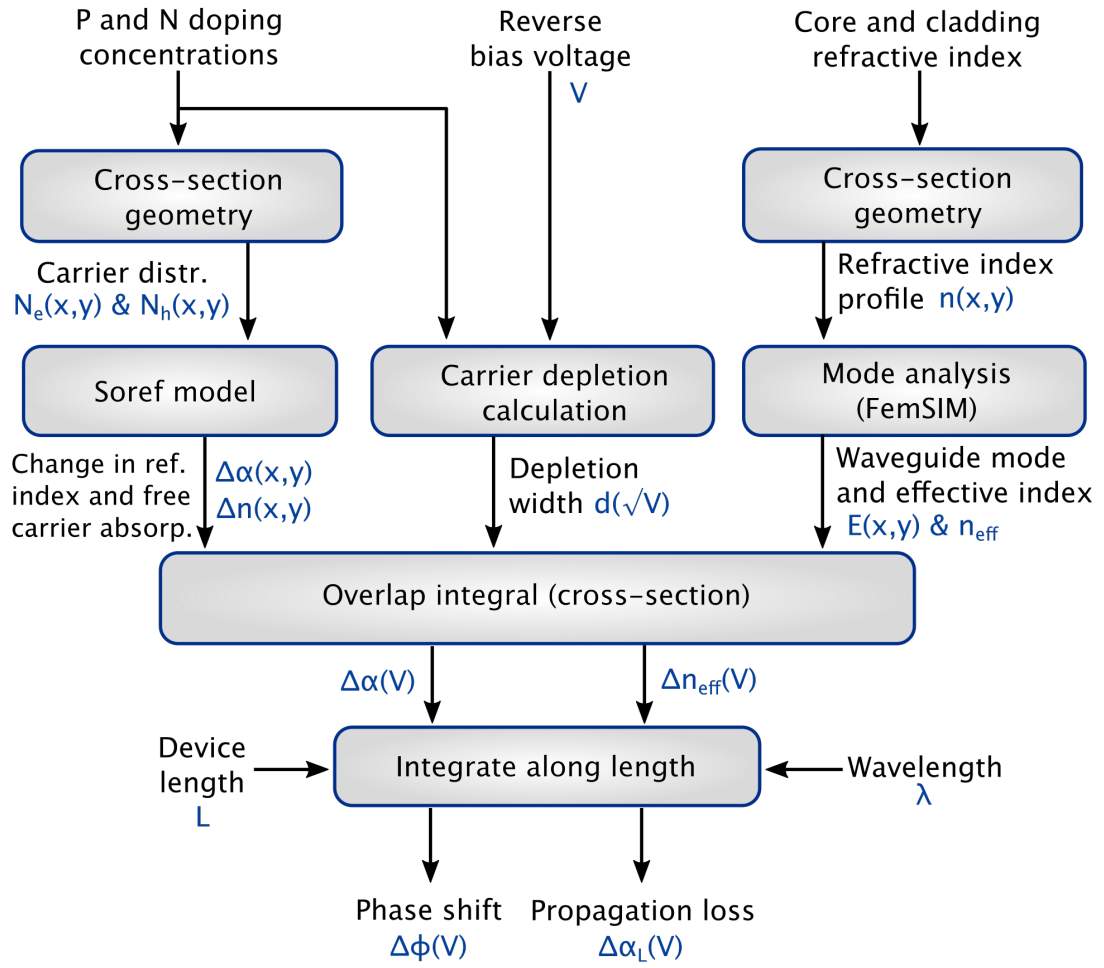


Figure 3.9: Flow chart depicting the steps used to obtain the simulated phase shift and absorption characteristics of the silicon depletion-mode phase shifter. Functions and modelling steps are shown in boxes, while inputs and outputs are shown in unboxed text.

effects. The method has been shown to be particularly accurate for lateral PN junctions, which are used in this thesis [118]. Figure 3.9 provides a summary of the steps involved in the phase shift and absorption modelling, noting the key parameters that are used and obtained along the way.

The main assumptions used in this model are that the total depletion width d is uniform along the waveguide height and modulator length, without any significant edge effects, and that it varies only in one direction when a voltage is applied, i.e., perpendicular to the plane of the PN junction. These assumptions are reasonable given that the junction is relatively tall. The optical mode profile is also assumed to be unaffected by introduction of doping, i.e., the carrier depletion calculation and mode analysis blocks in Figure 3.9 do not interact. This is valid as the change in refractive index from introduction of dopants is

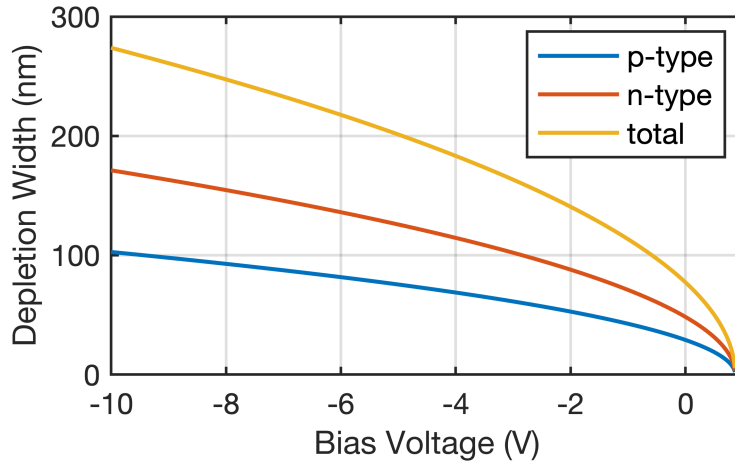


Figure 3.10: Depletion width shown according to the p-type region, n-type region and total, for doping levels $N_e = 3 \times 10^{17} \text{ cm}^{-3}$ and $N_h = 5 \times 10^{17} \text{ cm}^{-3}$.

very small compared to the index of the undoped waveguide, which is expected from Figure 2.3.

The carrier concentrations in the p-type and n-type regions of junction are assumed to be uniform and exactly equal to targeted doping levels used for each side. They are labelled as N_h and N_e for the concentration of free holes and electrons, respectively. In this case, the total depletion width of the PN junction can be described by summing the contributions of the two doping types through [121, 73]

$$\begin{aligned}
 d &= d_p + d_n = \sqrt{\frac{2\epsilon}{q_e} \frac{N_e (V_b - V)}{N_h (N_e + N_h)}} + \sqrt{\frac{2\epsilon}{q_e} \frac{N_h (V_b - V)}{N_e (N_e + N_h)}} \\
 &= \sqrt{\frac{2\epsilon}{q_e} \left(\frac{1}{N_h} + \frac{1}{N_e} \right) (V_b - V)},
 \end{aligned} \tag{3.5}$$

where d_p and d_n are the depletion widths in the p-type and n-type regions, respectively, ϵ is the permittivity of silicon, q_e is the electronic charge, V is the applied voltage, and V_b is the built-in voltage or junction potential, given by

$$V_b = \frac{kT}{q} \ln \left(\frac{N_h N_e}{N_i^2} \right). \tag{3.6}$$

Here, k is Boltzmann's constant, T is (room) temperature, and N_i is the intrinsic carrier concentration of silicon, which is taken as $1.45 \times 10^{10} \text{ cm}^{-3}$.

Although it is a textbook formula, it is worth plotting Equation 3.5 against voltage to see its shape as it has large impact on the characteristics of the phase

shift produced by the junction. To this end, Figure 3.10 shows the depletion width plotted against bias voltage for p- and n-type doping concentrations of $5 \times 10^{17} \text{ cm}^{-3}$ and $3 \times 10^{17} \text{ cm}^{-3}$, respectively. The square-root dependence of the depletion width on applied voltage is clearly visible. As would be expected, the built-in voltage occurs above 0 V, which shows that the depletion-mode phase shifter can still be considered to be in reverse bias for slightly positive values of voltage ($\lesssim 0.8 \text{ V}$).

Using this expression for the depletion width, and knowing the geometry of the rib waveguide, the carrier concentrations can then be expressed in terms of a cross-sectional distribution $N_e(x, y)$ and $N_h(x, y)$ for various values of applied voltage. The carrier concentrations are taken as zero within the depletion region, with a brick-wall transition from their doped levels. The Soref-Bennet model (Equations 2.19 and 2.20) is used to map the free-carrier distribution to corresponding cross sections of refractive index and absorption changes, labelled $\Delta n(x, y)$ and $\Delta \alpha(x, y)$, respectively.

From here, perturbation theory [122] can be applied to calculate the complex refractive index change (i.e., the change in effective index and absorption coefficient) with applied voltage. If the overall change in index from its unperturbed value is small (which is expected from Figure 2.3), then an overlap integral between the optical mode inside the waveguide and carrier distribution is sufficient to give the change in value. This overlap integral for effective index change with voltage is given as

$$\Delta n_{\text{eff}}(V) = \frac{\iint n(x, y) \cdot \Delta n(x, y) \cdot |E(x, y)|^2 \, dx \, dy}{n_{\text{eff}} \iint |E(x, y)|^2 \, dx \, dy}, \quad (3.7)$$

where $|E(x, y)|^2$ is the normalised optical intensity of the fundamental guided mode, and $n(x, y)$ is the refractive index cross section of the waveguide (a constant value of 3.48 inside the waveguide for silicon and 1.44 outside for silicon dioxide). The double integral is taken over the entire two-dimensional region of simulated optical mode. As $\Delta n(x, y)$ is zero outside the waveguide due to the absence of free carriers, the integral in the numerator of this expression is essentially taken over the cross-sectional area A of the waveguide, while the integral in the denominator remains over the entire field. The differentials dx and dy are the step size for numerical integration, and are determined by the step size of the output of the earlier optical mode simulation. A similar overlap integral

gives the absorption coefficient change with voltage as

$$\Delta\alpha(V) = \frac{\iint \Delta\alpha(x, y) \cdot |E(x, y)|^2 dx dy}{\iint |E(x, y)|^2 dx dy}. \quad (3.8)$$

Finally, the phase shift and change in insertion loss follow by integrating along the length of the phase shifter. As the junction is homogenous in this direction, a simple multiplication by the length L is sufficient. The phase shift then takes the form of $\Delta\phi = \frac{2\pi}{\lambda} \Delta n_{\text{eff}} L$ (as in Equation 2.4), and the change in insertion loss is given (in linear terms) by $\exp(\Delta\alpha L)$. It is important to note that, as the depletion width increases with greater reverse bias, the total insertion loss *decreases* by the amount of $\exp(\Delta\alpha L)$ from whatever value it began with, which is given by the sum of the unperturbed free carrier absorption, side-wall losses, coupling losses, etc.

3.3.2 Influence of Doping on Extinction Ratio, Insertion Loss and Level Mismatch Ratio

With a model established for the absorption and refractive index change produced by the depletion-mode phase shifters, analysis can be carried out on the effect of doping levels on the insertion loss, V_π and other characteristics of MZMs that use such devices.

Figure 3.11 shows the propagation loss resulting from free carrier absorption in a 1.5 mm MZM, plotted against bias voltage for various junction doping levels ranging from $0.5 \times 10^{17} \text{ cm}^{-3}$ to 10^{18} cm^{-3} . The waveguide geometry is taken as shown in Figure 2.4(b). The p-type and n-type doping levels are assumed to be equal, and the junction is placed in the centre of the waveguide. A full optimisation of the junction can be carried out by using imbalanced doping levels and offset in the junction placement [78], which for simplicity is not investigated here. The propagation loss is largest for each doping level at the smallest reverse bias voltage, and decreases with increasing reverse bias. As expected, the loss rises with increased doping levels due to greater free-carrier absorption, with loss values ranging from much less than 0.5 dB to greater than 3 dB. It should be emphasised that, even in the case of an ideal SiP MZM with no other sources of propagation loss or coupling loss, the operation at quadrature point will increase the insertion loss by 3 dB from the values shown in Figure 3.11.

Figure 3.12 shows the phase shift produced by the same MZM conditions. The

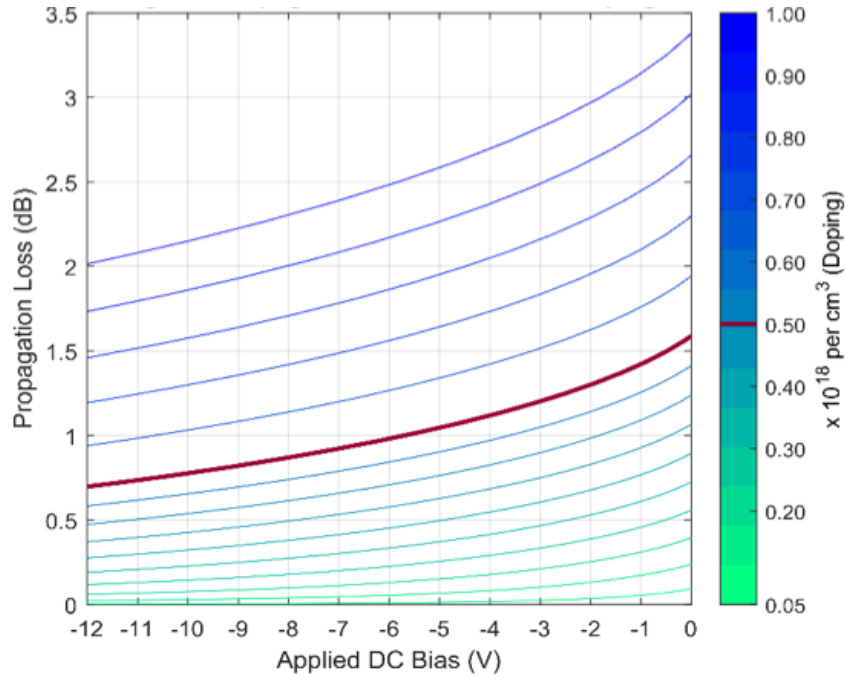


Figure 3.11: Propagation loss due to free-carrier absorption, plotted against bias voltage for various doping levels in an MZM with 1.5 mm active length.

non-linear nature of the phase-voltage relationship is clear, with greater phase shift achieved per volt at smaller reverse biases. To help show how the phase would vary in the presence of an applied RF voltage, the phase shifts are all plotted relative to an reference bias point of -4 V. The amount of phase shift achieved for a given voltage swing depends heavily on the doping level of the junction. For example, the lowest doping level shown ($0.5 \times 10^{17} \text{ cm}^{-3}$) provides less than 20° phase shift across the entire 12 V sweep as it flatlines quickly with increased voltage, rendering it an unsuitable choice despite its extremely low level of free-carrier absorption indicated by Figure 3.11. Heavier doping naturally increases the phase shift achieved per volt, at the expense of higher insertion loss. A moderate level of doping ($5 \times 10^{17} \text{ cm}^{-2}$), chosen as it is close to the measured values for the experimental SiP MZM in our lab, is highlighted in red as an example. As shown by the annotations on the graphs, to achieve a 90° phase shift in one arm (suitable for full extinction in push-pull dual-drive operation), a peak-to-peak voltage swing of 6 V is required, totalling to a combined peak-to-peak RF voltage of 12 V in the two arms. Such a large voltage is potentially prohibitive for use in data-centre applications. However, full extinction is not always required in these scenarios; a lower voltage, providing extinction ratios in the region of 6 dB or less is often sufficient.

To determine the extinction ratio and other system parameters, the phase shift

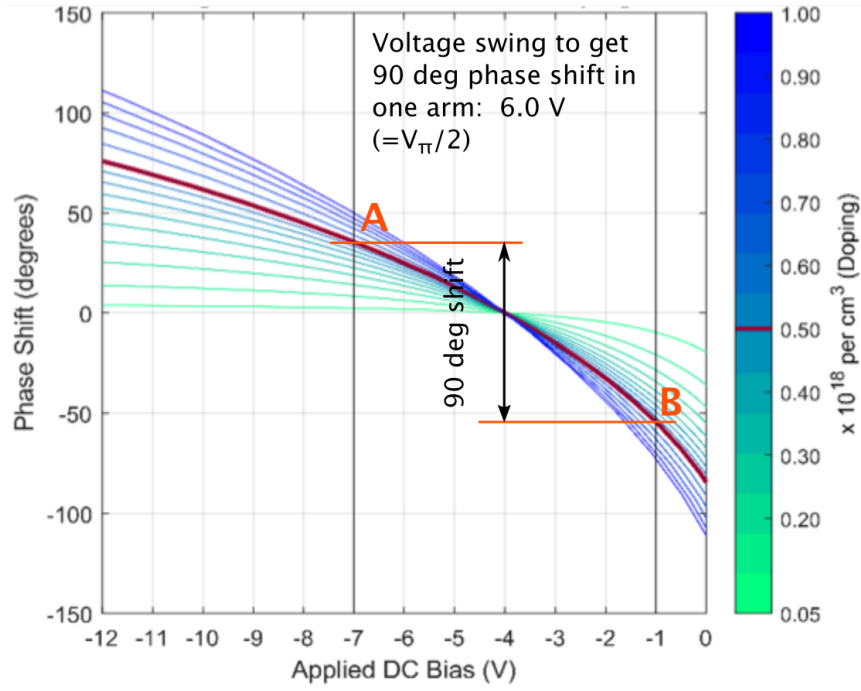


Figure 3.12: Phase shift in a single MZM arm, plotted against bias voltage for various doping levels in an MZM with 1.5 mm active length. The phase shift is shown relative to a reference bias point of -4 V.

and absorption must be included appropriately into a model for the electric field E_{out} at the output of the MZM structure under the influence of two drive voltages $V_1(t)$ and $V_2(t)$. This is described below as a more complete form of the simplified expression that was used in Equation 2.2. This output field can be described in terms of the sum of the electric fields in the upper (E_1) and lower (E_2) arms of the MZM as $E_{out} = E_1 + E_2$, where

$$E_1(t) = \frac{E_{in}}{\sqrt{2}} \cdot \left(\sqrt{\frac{1}{2} + \epsilon} \right) \cdot A_1(t) \cdot \exp \left(i\phi_{offset} + i\phi_1(t) \right), \quad (3.9)$$

$$E_2(t) = \frac{E_{in}}{\sqrt{2}} \cdot \left(\sqrt{\frac{1}{2} - \epsilon} \right) \cdot A_2(t) \cdot \exp \left(i\phi_2(t) \right). \quad (3.10)$$

Here, E_{in} is the electric field at the input of the MZM. The term ϵ represents the difference in the upper and lower power split ratios, which will be assumed to be zero throughout this thesis, i.e., the light splits equally between the two MZM paths. The phase term ϕ_{offset} is the phase offset introduced by length asymmetry or thermal means to allow tuning of the phase without adjusting electrical bias. The terms $A_1(t) = \sqrt{\exp(\alpha(V_1(t)) \cdot L)}$ and $A_2(t) = \sqrt{\exp(\alpha(V_2(t)) \cdot L)}$ account for the propagation loss of the electric field in the waveguide, which is here due to free-carrier absorption (α), while the terms $\phi_1(t) = \frac{2\pi}{\lambda} \Delta n_{eff}(V_1(t))L$ and

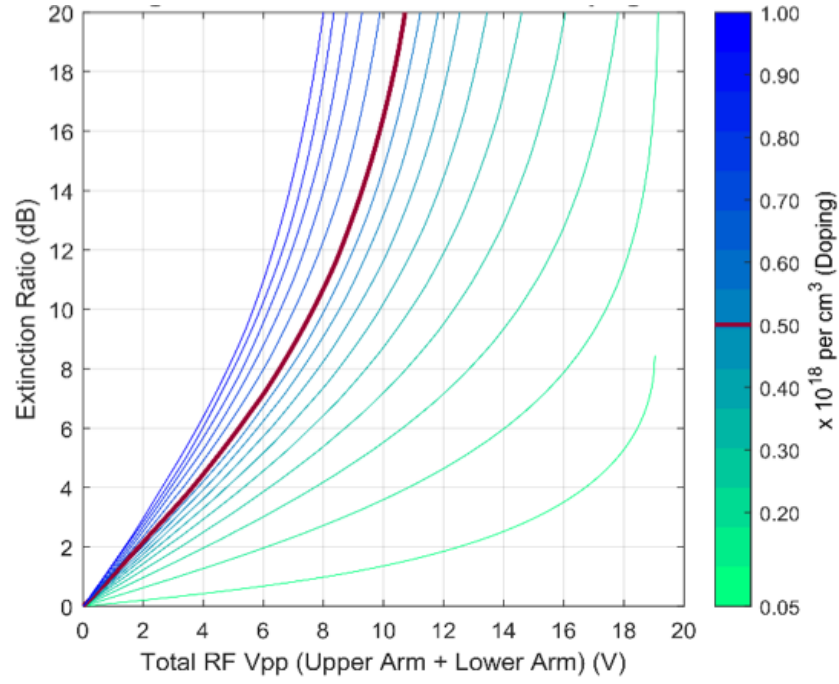


Figure 3.13: Extinction ratio against total peak-to-peak RF voltage in a dual-drive push-pull 1.5 mm MZM operating a quadrature, with applied bias of -4 V, for various doping levels.

$\phi_2(t) = \frac{2\pi}{\lambda} \Delta n_{eff}(V_2(t))L$ represent the phase shift from refractive index change in the usual fashion.

The extinction ratio achieved for a given RF voltage depends on the bias point chosen. Figure 3.13 plots the extinction ratio against the total peak-to-peak RF voltage in the two arms for the various doping levels, while both arms have the same bias of -4 V applied. The MZM is operated at quadrature by setting ϕ_{offset} to $\pi/2$. The accessible range of RF voltage at this bias point is bounded by the need to keep the MZM in depletion mode, i.e., below the phase shifters' built-in voltage of approximately 0.7 V. Hence the total peak-to-peak voltage swept does not exceed 19 V (9.5 V in each arm). The voltage required for full extinction can be interpreted from such a graph by seeing where the curves become (asymptotically) vertical, although to keep this graph legible only extinction ratios of up to 20 dB are shown. Other desired values of ER can be read from the y-axis as needed. As annotated in the graph, to reach a 6 dB ER with the example doping of $5 \times 10^{17} \text{ cm}^{-3}$, a total peak-to-peak RF voltage of 5.2 V is required (2.6 V in each arm).

When silicon MZMs are used for multilevel modulation formats such as PAM4, the non-linear phase-voltage relationship presents a potential issue. It can be

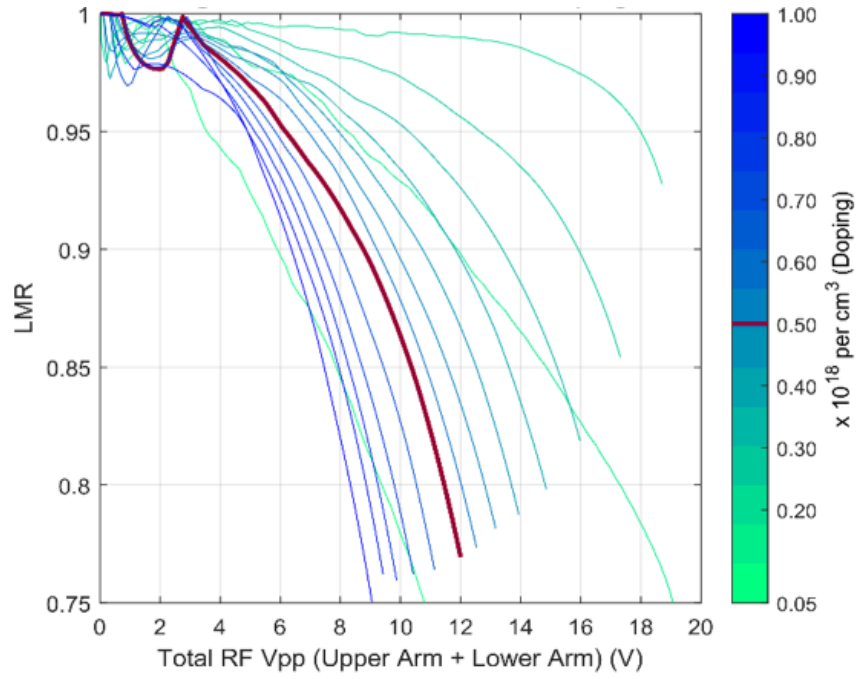


Figure 3.14: Level mismatch ratio against total peak-to-peak RF voltage in a dual-drive push-pull 1.5 mm MZM operating a quadrature, with applied bias of -4 V, for various doping levels.

seen from Figure 3.12 that the phase swings produced are not symmetric around a given bias point. For the reference bias of -4 V, for example, the phase shift to the symbol rail A is approximately 45° , while the the corresponding shift to the symbol rail B is 55° . In addition, the change in absorption with different voltage levels in the phase shifters will affect the output amplitude of the MZM unevenly, as each arm will be at a high or low voltage respectively during push-pull operation. These two effects can influence a PAM4 signal by distorting the spacing of its symbol levels. One metric to describe this distortion is the level separation mismatch ratio (LMR), as defined in IEEE P802.3bj clause 94.t. For a PAM format with M symbol levels, this is calculated as

$$LMR = \frac{(M-1)S_{min}}{OMA}, \quad (3.11)$$

where S_{min} is the smallest difference between the symbol power levels. The term in the denominator is the optical modulation amplitude (OMA), defined as the difference in optical power between the highest and lowest symbol levels in the signal. Figure 3.13 shows the LMR against the total-peak-to-peak RF voltages for the same MZM parameters as before, with a PAM4 signal applied. As could be expected, the LMR degrades as the RF voltage increases, with heavier doping levels being more strongly affected due to the larger phase shift produced by

the junctions. The LMR in the two lowest dopings simulated does not follow the trend of the other curves as the overall swing for these dopings is very low, affecting the calculation of the amplitude levels in the signal. The fluctuations present in the LMR curves for all the dopings at low voltages is similarly due to difference in amplitude levels being small for these swings. Although the LMR clearly does degrade, the overall drop is quite small, particularly for reasonable voltage levels. For example, for the total peak-to-peak voltage of 5.2 V that was needed to produce a 6 dB ER in the highlighted doping, the LMR is greater than 0.95.

3.3.3 Modelling the RC Bandwidth of SiP MZMs

The highest limit on the modulation bandwidth of a silicon phase shifter (and thus the MZM structure) is given by its intrinsic RC bandwidth, i.e., assuming that any driver used has an ideal output impedance of $0\ \Omega$ and so does not influence the device speed. The capacitance C of the PN junction can be modelled after a textbook parallel plate capacitor as

$$C = \epsilon A/d, \quad (3.12)$$

where ϵ is the permittivity of silicon, A is the area of the capacitor, and d is the depletion width of the junction as described by Equation 3.5. Here, the area A is given by the by the product of the rib waveguide height h_{rib} and the active length of the phase shifter L_{mzm} . As the total capacitance is linearly proportional to length, an often-used quantity is the linear capacitance C_l , defined by $C = C_l L_{mzm}$. The resistance R of the phase shifter depends on the doping and dimensions of its various sections that were depicted in Figure 3.8. The sections must have their resistance calculated separately and then summed, with each following

$$R = \rho W/A, \quad (3.13)$$

where ρ is the resistivity of the doped silicon, W is the width of the section, and A is the area as defined above. As total resistance is inversely linearly proportional to the active length, the linear resistance R_l is defined by $R = R_l/L_{mzm}$.

The resistivities of p-type and n-type silicon are determined respectively from

$$\rho_n = 1/N_e q_e \mu_e, \quad (3.14)$$

$$\rho_p = 1/N_h q_h \mu_h, \quad (3.15)$$

Table 3.1: Fit parameters for boron and phosphorous to be used in the Caughey-Thomas model.

	Boron	Phosphorous
$\mu_{min} \left[\text{cm}^2/\text{Vs} \right]$	44.9	68.5
$\mu_{max} \left[\text{cm}^2/\text{Vs} \right]$	470.5	1414
$N_r \left[\text{cm}^{-3} \right]$	2.23×10^{17}	9.20×10^{16}
$\alpha \left[\right]$	0.719	0.711

where N_e and N_h are the respective free electron and hole concentrations, and μ_e and μ_h are the respective mobilities of those electron and holes. Carrier mobilities depend on the level of carrier concentration present and the type of dopant used (typically boron for p-type silicon and phosphorus for n-type silicon). Here, we use the empirical Caughey-Thomas model for the mobilities [123], which has been shown to fit very well against experimental data for carrier concentrations in the range of 10^{15} cm^{-3} to 10^{20} cm^{-3} . The model follows a modified hyperbolic tangent function in truncated continued fraction form as

$$\mu = \frac{\mu_{max} - \mu_{min}}{1 + (N/N_{ref})^\alpha} + \mu_{min}, \quad (3.16)$$

where μ is mobility, N is carrier concentration, and μ_{max} , μ_{min} , N_{ref} and α are fit parameters that depend on the choice of dopant used. Here, we take fit parameters for boron and phosphorous from [121], shown in Table 3.1. The resulting mobility and resistivity for boron-doped and phosphorous-doped silicon are shown against carrier concentration in Figures 3.15(a) and 3.15(b), respectively. At low doping concentration, the mobility is primarily limited by phonon scattering and is almost constant, but decreases due to greater incidence of impurity scattering as the doping level is increased. Despite this decrease in carrier mobility, the resistivity still drops with increasing carrier concentration.

With this model for the resistance and capacitance of the phase shifter, the 3 dB bandwidth can then be determined from $f_{3\text{dB}} = 1/2\pi RC$. In the calculation for resistance, the contributions of the heavily doped ($\sim 10^{20} \text{ cm}^{-3}$) p++ and n++ contact regions can generally be omitted as their resistivity is two orders of magnitude lower than the more lightly doped ($\sim 10^{17} \text{ cm}^{-3}$) p and n regions of the central waveguide. Figures 3.16(a) to 3.16(c) show, respectively, the capacitance, resistance and intrinsic 3 dB bandwidth of a typical depletion-mode phase shifter against the central p-n doping level, for various active region lengths. The applied

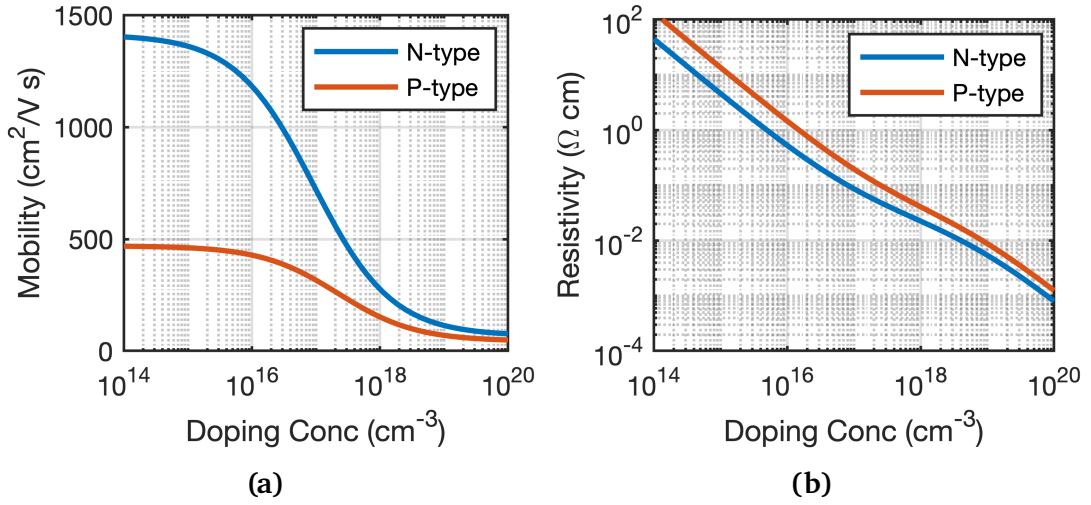


Figure 3.15: (a) Carrier mobility and (b) material resistivity against carrier concentration for silicon doped with boron (p-type) and phosphorus (n-type).

bias voltage is -1 V . As expected from their dependence on cross-sectional area, the capacitance increases with length, while the resistance correspondingly decreases. Notably, the intrinsic RC bandwidth is independent of length, due to the respective length dependence of R and C cancelling in Equation 2.22:

$$f_{3\text{dB}} = \frac{1}{2\pi(R_l/L_{\text{mzm}})(C_l L_{\text{mzm}})} = \frac{1}{2\pi R_l C_l}. \quad (3.17)$$

It should be also noted that the intrinsic RC bandwidth increases with greater doping concentration, due to the resistance decreasing faster than the capacitance increases.

As the optical field travels mostly in the central waveguide region, it is possible to increase the doping levels in part of the slab waveguide (P/N+ in Figure 3.8) to decrease the overall resistance without affecting the optical absorption, as shown in Figure 3.17(a) for an intermediate doping level of $2 \times 10^{18} \text{ cm}^{-3}$. The capacitance remains unchanged, as it is determined by the carriers present immediately around the depletion region in the waveguide centre. The addition on this intermediate doping region significantly improves the intrinsic RC bandwidth, as shown in Figure 3.17(b). It does also cause the intrinsic bandwidth to now decrease rather than increase with greater junction doping concentration. This is due to the intermediate doping causing the total resistance to follow $X + 1/N$, where X is a constant offset and N is doping, rather than the simple $1/N$ dependence of before. Due to capacitance following $\sqrt{N}$ via the depletion width, the decrease in resistance is no longer enough to

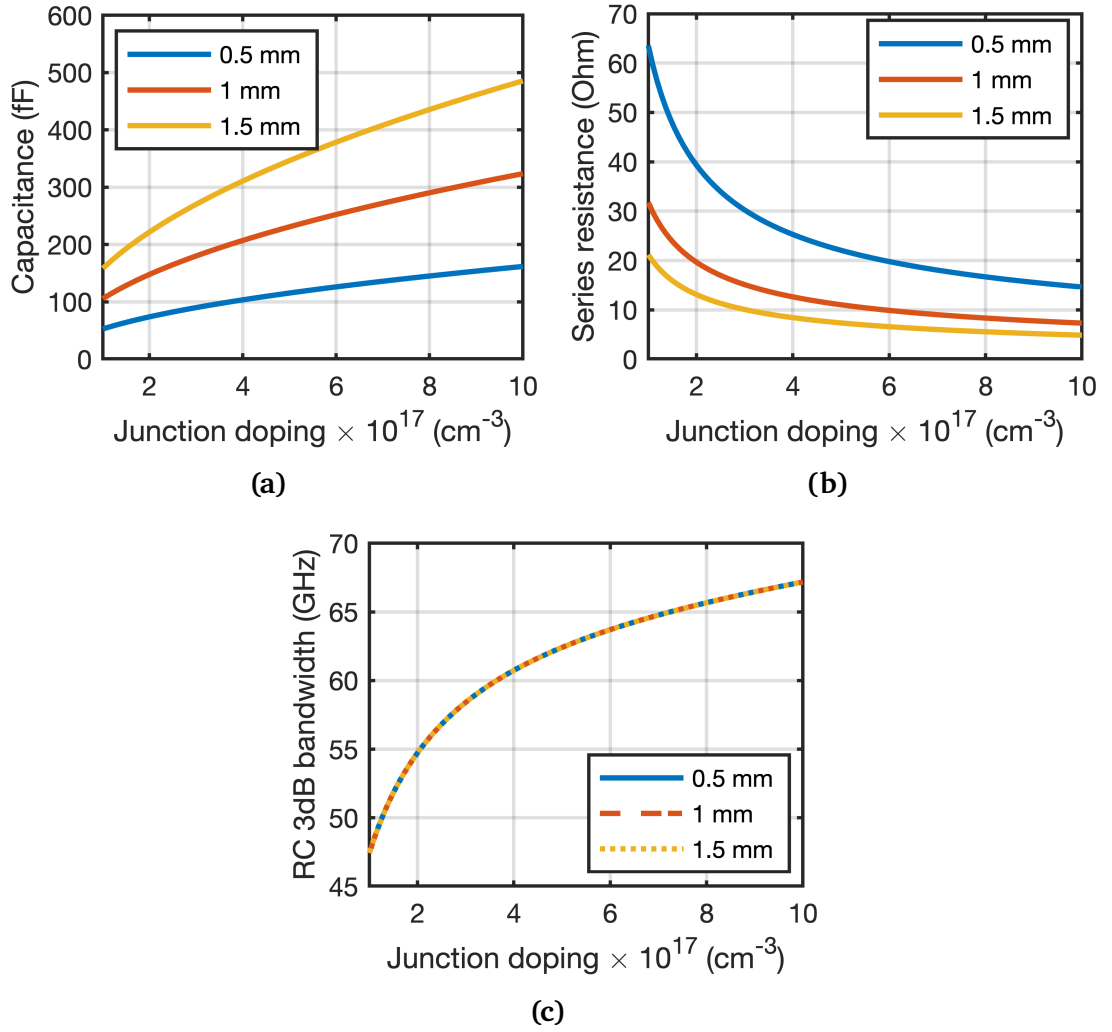


Figure 3.16: (a) Simulated capacitance, (b) resistance and (c) intrinsic RC 3 dB bandwidth of a silicon phase shifter at -1 V bias, plotted against junction doping for various active region lengths.

win over the increase in capacitance at higher doping.

This relationship can be seen more clearly in Figure 3.17(c), where the intrinsic RC bandwidth is plotted over a wider central junction doping range, for various levels of intermediate doping. (No values are plotted below approximately $2.8 \times 10^{16} \text{ cm}^{-3}$, the junction doping value at which the depletion region extends beyond the rib waveguide width and, hence, beyond the range of relevance of our model.) This graph also suggests that the highest possible intermediate doping levels should be used to maximise bandwidth. However, in reality, there is a limit to the doping levels that can be practically used to avoid diffusion of the intermediate dopants into the central waveguide during the annealing process, which has been shown to cause significant optical attenuation [124]. In

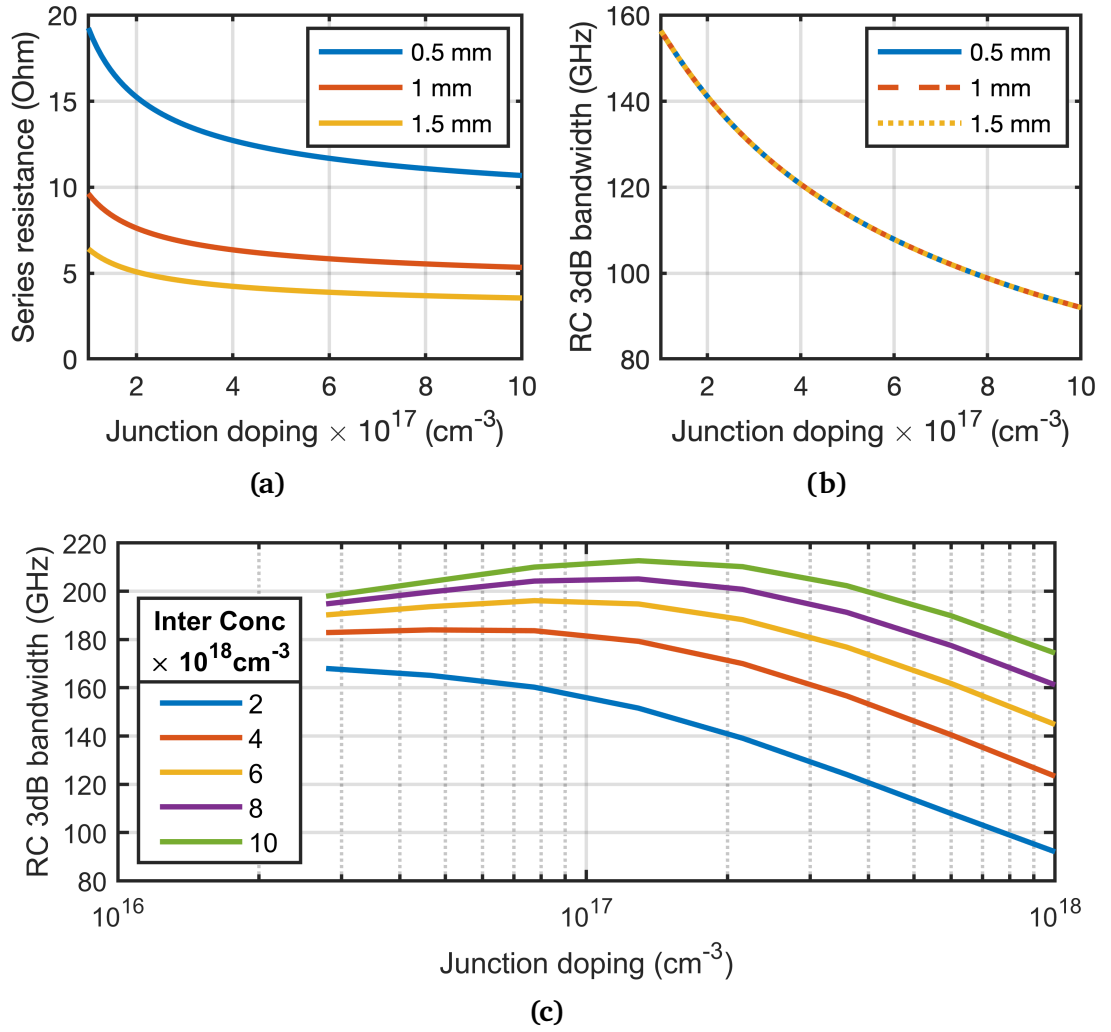


Figure 3.17: (a) Simulated phase shifter resistance and (b) intrinsic 3 dB bandwidth against central junction doping, with intermediate doping of $2 \times 10^{18} \text{ cm}^{-3}$. (c) The intrinsic bandwidth over a wider range of junction doping, showing the effect of intermediate doping levels.

practice, intermediate doping levels of $2 \times 10^{18} \text{ cm}^{-3}$ are typical [125, 126, 100], although heavier doping may be used in certain designs if placed suitably far from the central waveguide [127].

It is clear from Figures 3.16(c), 3.17(b) and 3.17(c) that intrinsic RC bandwidths of SiP phase shifters can be many tens or even hundreds of GHz. However, in reality, the RC bandwidth is heavily reduced by the output impedance of electrical drivers, particularly in a typical 50Ω environment. Assuming this impedance is purely resistive, and labelling it R_{dr} , the 3 dB RC bandwidth will now be given

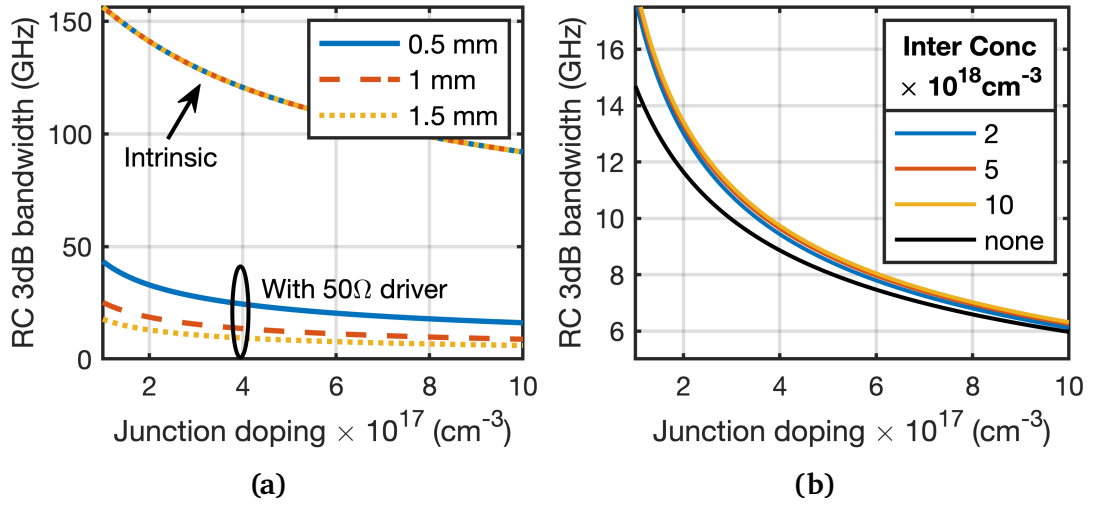


Figure 3.18: (a) 3 dB RC bandwidth for various lengths of MZMs, with intermediate doping of $2 \times 10^{18} \text{ cm}^{-3}$, with and without a 50Ω driver. (b) The bandwidth for the case of the 1.5 mm MZM with a 50Ω driver, for various intermediate dopings.

by

$$f_{3\text{dB}} = \frac{1}{2\pi(R_l/L_{mzm} + R_{dr})(C_l L_{mzm})} = \frac{1}{2\pi(R_l C_l + R_{dr} C_l L_{mzm})}. \quad (3.18)$$

This expression shows that the bandwidth will reduce with larger driver impedance, and that it also now acquires an inverse linear dependence on the length of the active region. This is demonstrated in Figure 3.18(a), where a 50Ω driver is added to an MZM with intermediate doping of $2 \times 10^{18} \text{ cm}^{-3}$, and the resulting bandwidth is plotted against central doping concentration. The figure clearly shows that the addition of the driver reduces the overall bandwidth by tens of GHz, with the longer MZMs being the worst affected. In particular, we note that higher junction doping levels for the 1.5 mm MZM result in very low bandwidths, below 10 GHz, emphasising that low junction doping levels may be required at that length.

It is worth taking this opportunity to also replot the 3 dB RC bandwidth of the MZM for various intermediate doping concentrations once the effect of the 50Ω driver is taken into account. This is shown in Figure 3.18(b) for a 1.5 mm MZM. The “none” line corresponds to the case where no additional intermediate doping is used, i.e., the junction doping level is used throughout the slab. Clearly, the addition of the driver reduces the heavy dependence of the bandwidth on the intermediate doping level, although it is still beneficial to have some level

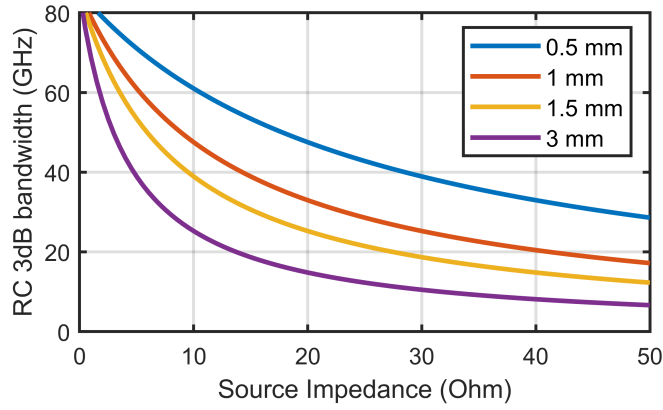


Figure 3.19: RC 3 dB bandwidth against driver output impedance, for MZMs with intermediate and central junction doping fixed at $2 \times 10^{18} \text{ cm}^{-3}$ and $2 \times 10^{17} \text{ cm}^{-3}$, respectively, for various active region lengths

of intermediate doping, particularly at lower levels of junction doping. This simplifies the model for the doping profile, allowing us to “set and forget” an intermediate concentration of $2 \times 10^{18} \text{ cm}^{-3}$ as in the previously mentioned papers.

In Figure 3.19, the 3 dB RC bandwidth is swept against driver output impedance, for MZMs with intermediate and central junction dopings fixed at $2 \times 10^{18} \text{ cm}^{-3}$ and $2 \times 10^{17} \text{ cm}^{-3}$, respectively, for various active region lengths. The bandwidth clearly drops with greater source impedance, with longer MZMs being more heavily affected. We note that the longest MZM shown (3 mm) has a bandwidth below 10 GHz, which suggests that SiP MZMs of such lengths cannot be used for very high-speed modulation. However, as discussed in Section 2.4, such lengths can be reached successfully provided that a travelling-wave design is used for the MZM electrodes. In addition, this analysis suggests that the driver impedance drastically affects the lumped RC bandwidth, showing it is beneficial to reduce this source impedance as much as possible, provided the driver can be freed from the constraint of a 50Ω environment. However, this requires careful driver design in order to maintain a low power consumption, as reducing the driver output impedance typically requires the output stage to have larger transistor size [64]. The performance of one such low-impedance driver design is discussed and tested experimentally in Sections 3.4 and 3.5. In addition, a more thorough description of the frequency response of lumped MZMs operated at high speeds will be given in Section 4.5, where the impact of circuit parasitics and the transit time of the optical wave through the phase shifter will be included in the frequency analysis for the discussion of signal equalisation.

3.4 Switched Capacitor Drive Scheme

As just discussed in Section 3.3.3, the bandwidth of modulators can be enhanced by reducing the output impedance of the driver used. This analysis focussed only on the resistive part of the driver impedance. In fact, the bandwidth can be enhanced even further by considering a capacitive drive scheme.

An additional capacitor can be placed (as part of the driver) in series with the phase shifter, as shown by the equivalent circuits in Figure 3.20. The benefit of this can be seen by considering the summation of two capacitors C_1 and C_2 in series. Their total capacitance C_T is given by $C_T = \frac{C_1 C_2}{C_1 + C_2}$. It is then clear that if $C_1 = C_2$, the overall capacitance will be halved. In our case, the additional series capacitor in the driver is designed to match the capacitance of the phase shifters in the MZM arms. The effect of this change to the driver is shown in Figure 3.21, where the normalised frequency response of the MZM arm (calculated from Equation 2.21) is compared between the $50\ \Omega$ drive scheme and the low-impedance $5\ \Omega$ drive scheme with and without the series capacitor. The 3 dB bandwidth is greatly enhanced by the low output impedance driver and series capacitor, rising to 35.4 GHz in comparison with 4.4 GHz from the $50\ \Omega$ drive case. There is a drawback to this scheme: as the voltage from the driver now falls across two capacitors in series, it is split in two according to the ratio of their capacitances. In this case, the voltage across the MZM arm will be halved. We do note, however, that the same effect happens in the case of the TW-MZM, where the voltage is divided equally across the output impedance of the driver and the $50\ \Omega$ termination resistor at the end of the MZM. As such, this capacitive drive scheme matches the voltage efficiency of the TW-MZM, while achieving high bandwidth and being freed from strict impedance matching conditions of the transmission line scenario.

3.4.1 Switched-Capacitor Drive Scheme for PAM4

A key benefit of the capacitive drive scheme is that it can be readily extended to multilevel modulation formats such as PAM4 or PAM8 by dividing the driver's series capacitor appropriately into a bank of parallel capacitors (which are still in series with the phase shifter). The required symbol amplitude level is then reached by switching on the drive voltage to all, some or none of the capacitors as needed. This provides a convenient way of generating multilevel PAM signals without the need for a potentially power-hungry electrical DAC. Such a “switched-

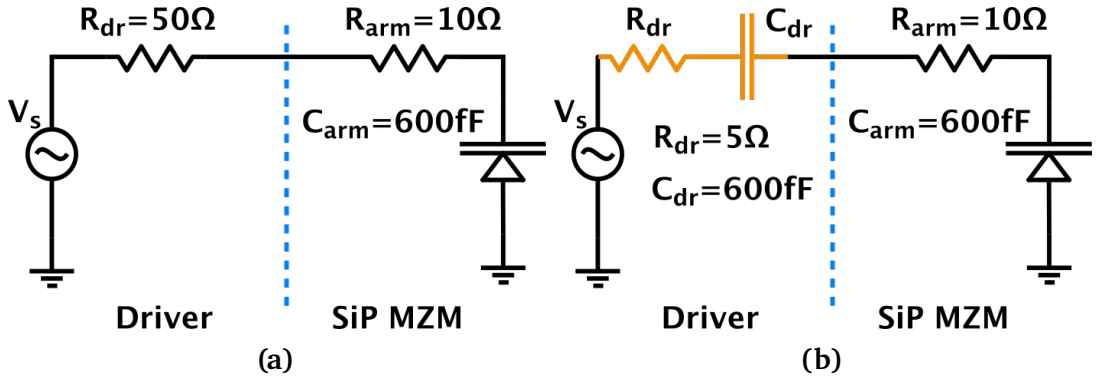


Figure 3.20: (a) Typical 50 Ω drive scheme for a silicon phase shifter. (b) Capacitive drive scheme, where the driver output impedance is reduced and a series capacitor is added to the driver..

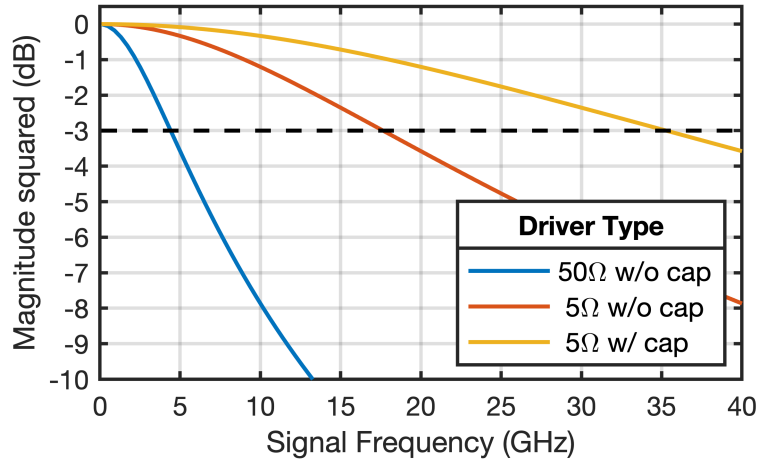


Figure 3.21: RC frequency responses for three different lumped drive schemes: 50 Ω , 5 Ω , and 5 Ω with series capacitor.

capacitor” driver was successfully designed and fabricated for a silicon photonic MZM in our research group during the course of another PhD project. Full details are given in the relevant thesis [128] and in published papers [129, 84]. The research work in my own PhD involved the experimental testing and optimisation of this driver circuit, thus the key details of the design are given below.

Figure 3.22 shows how a switched capacitor bank is used to generate a PAM4 signal. Firstly, the driver applies a relevant DC bias to the phase shifter in a way that isolates it from the high-speed RF circuitry. For this purpose, a large-value 120 k Ω on-chip resistor can be used rather than an on-chip inductor, as the switched-capacitor bank naturally acts as the capacitive part of a bias-tee. This conveniently saves on chip area. For the RF signal, three binary logic signals X , Y and Z control three respective RF voltage sources, V_x , V_y and V_z . Each of

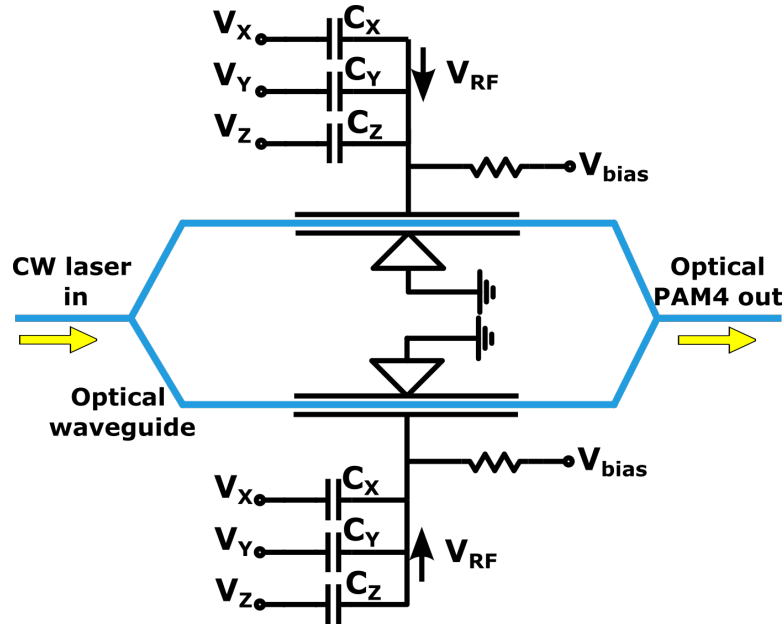


Figure 3.22: Switched-capacitor, dual-drive, push-pull, PAM4 drive scheme for the silicon MZM.

these voltages then powers one of three capacitors in the bank, which are labelled C_x , C_y and C_z , respectively. The resulting signal from this capacitor bank, which we will call V_{pam} , then falls across the silicon phase shifter in the MZM arm. Due to the charge balance principle, which demands that the average capacitor current over one switching interval equals zero, the total charge $C_{arm}V_{pam}$ on the phase shifter capacitor can be expressed as

$$C_{arm}V_{pam} = C_x(XV_x - V_{pam}) + C_y(YV_y - V_{pam}) + C_z(ZV_z - V_{pam}). \quad (3.19)$$

Rearranging this equation, the voltage across the phase shifter follows as

$$V_{pam} = \frac{C_x V_x X + C_y V_y Y + C_z V_z Z}{C_x + C_y + C_z + C_{arm}}. \quad (3.20)$$

It is clear that by switching the 0/1 logic values, voltages for four symbol levels can be created across the phase shifter, with the selected level depending on how many of the logic streams X , Y and Z are in an “on” state. For dual-drive, push-pull operation, the logical inverse of X , Y and Z are applied to the other arm of the MZM using a second set of capacitors.

The logic signals X , Y and Z are derived originally in the circuit from two input NRZ signals (A and B) such that $X = A \oplus B$, $Y = A$, and $Z = A \oplus \bar{B}$. This logic includes a Gray coding scheme to reduce the number of bit transitions and thus

Table 3.2: Logic table for the four PAM levels using the switched capacitor bank.

Level	A	B	X	Y	Z	V_{pam}
One	0	0	0	0	0	0
Two	0	1	1	0	0	$1/3V_d$
Three	1	1	1	1	0	$2/3V_d$
Four	1	0	1	1	1	V_d

improve the signal integrity. The required logic states to achieve each of the PAM4 levels are shown in Table 3.2. The actual value of V_{pam} for each level depends on the values of the relevant capacitors in the bank and the drive voltage they experience. Assuming that $V_x = V_y = V_z$ and that each of C_x , C_y and C_z equal one third the value of C_{arm} , then the four voltage levels will be equally spaced. However, as an MZM has a sinusoidal transfer function, the capacitors should be resized appropriately to generate the correct voltage levels for maximum level mismatch ratio (LMR).

As the peak-to-peak outputs of transistors in the 65 nm CMOS process used for the driver are limited to approximately 1.2 V, additional level shifter and voltage doubler circuits are required before the capacitor banks in order to achieve the targeted 2.4 V open-circuit voltage. This design step is necessary to achieve a 1.2 V drive swing across the capacitance of the phase shifter. The circuit also contains high-speed D-type flip-flops between the Gray-coding circuitry and the level shifter in order to correctly synchronise the logic streams, as the Gray coder's OR gate, AND gate, and buffer gate all differ in propagation delay.

3.4.2 Optimisation of Bias and Doping Levels for Switched-Capacitor Lumped MZMs

The switched-capacitor driver design can clearly extend the RC bandwidth of lumped MZMs into the tens of GHz, and facilitates the use of multilevel modulation formats through its capacitor bank. However, at these modulation frequencies, the bandwidth is also limited by travelling wave effects (RF reflections, impedance mismatch, and the velocity mismatch between the optical and electrical signals in the device). Hence, to minimise this, the effective MZM length is constrained to below 2 mm, often much less if the modulation frequency is very high. Unfortunately, the peak-to-peak RF voltage swings provided by current CMOS technology are in the order of 1 V to 2 V,

unless significant RF amplification is used in the driver design. As discussed in Section 3.4.1, the switched-capacitor driver circuit in this work is limited to a 1.2V swing across an MZM arm. This is insufficient to drive short SiP depletion-mode MZMs through a full half-wave voltage (V_π), leading to low signal extinction ratios (ER). The V_π can be reduced by increasing the doping levels of the SiP phase shifters, but this leads to increased insertion loss (IL) in the device. Unlike traditional LiNbO₃ MZMs, the choice of bias point affects both the ER and IL of SiP MZMs, as the phase shift efficiency and free-carrier absorption of the phase shifters vary with voltage. To compound this, as the phase shifters respond non-linearly to voltage, advanced modulation formats such as PAM4 and PAM8 can suffer from mismatched symbol level spacing. In this section, these issues are taken into account, optimising the device bias point and doping levels by defining a transmission link penalty (TLP) as a suitable figure of merit.

The SiP phase shifters simulated here are based on [78], with a lateral PN junction and balanced p-type and n-type doping levels. The junction is set in centre of the waveguide to mimic the design of phase shifters used in the experimental switched-capacitor MZM device. Figure 3.23 shows the optical mode confinement, junction dimensions, doping profile and intrinsic depletion region. As described in Section 3.3.1, the doping is assumed to be distributed uniformly in the p-type and n-type regions, and the depletion region is taken to be perfectly vertical.

The phase shift and free-carrier absorption are simulated for various doping levels, and are shown against applied voltage in Figure 3.24 for a phase shifter of length 1 mm. The chosen doping levels are $2 \times 10^{17} \text{ cm}^{-3}$, $5 \times 10^{17} \text{ cm}^{-3}$ and

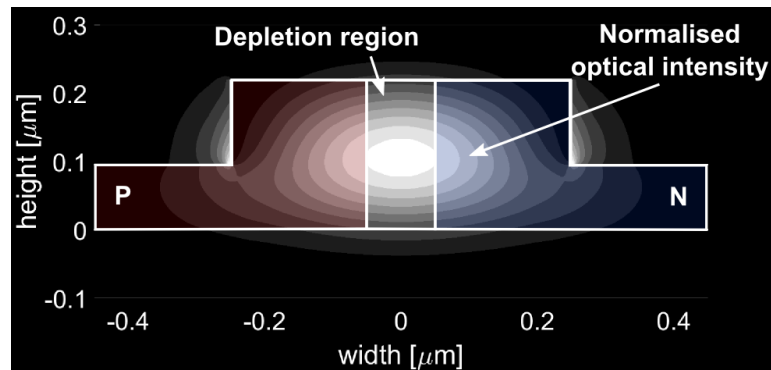


Figure 3.23: Cross section of the SiP rib-waveguide phase shifter, showing the optical mode, the p-type and n-type doping regions, and the depletion region.

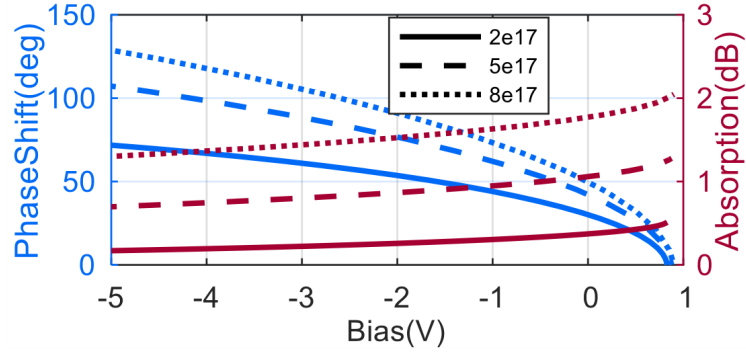


Figure 3.24: Phase shift and absorption against bias voltage for various levels of doping in a 1 mm SiP phase shifter.

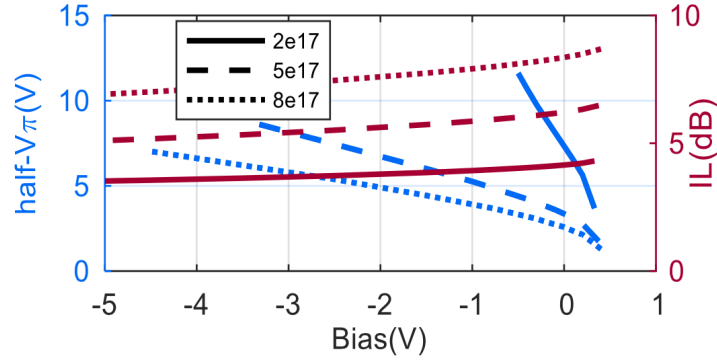


Figure 3.25: Half- V_{π} and insertion loss against bias for various levels of doping in a silicon MZM with 1.5 mm phase shifters.

$8 \times 10^{17} \text{ cm}^{-3}$, and can be considered as typically low, medium and high levels of doping, respectively. It can be seen that the obtained phase shift per volt is maximum at low biases, i.e., as close as possible to the built-in voltage V_b . For a given V_{pp} , this bias condition is met at $V_b - \frac{V_{pp}}{2}$. Unfortunately, this also maximises the free-carrier absorption. To illustrate the effect this trade-off has in an MZM structure, the resulting half- V_{π} (chosen as the device is designed for dual-drive push-pull configuration) and IL are plotted against bias in Figure 3.25 for a 1.5 mm MZM, assuming no additional coupling or propagation losses. It should be noted that for optimal performance, the MZM is operated at its quadrature point, with the $\frac{\pi}{2}$ phase shift reached here by a fixed path difference in the MZM arms so as not to disturb the bias of the phase shifters. The graph shows that, for all the chosen doping levels, a small decrease in IL is obtained at greater reverse biases (approximately 1 dB and 2 dB for the low- and high-doped cases, respectively, as the bias moves from V_b to -5 V). However, in all cases, the corresponding half- V_{π} increases significantly with reverse bias, which will

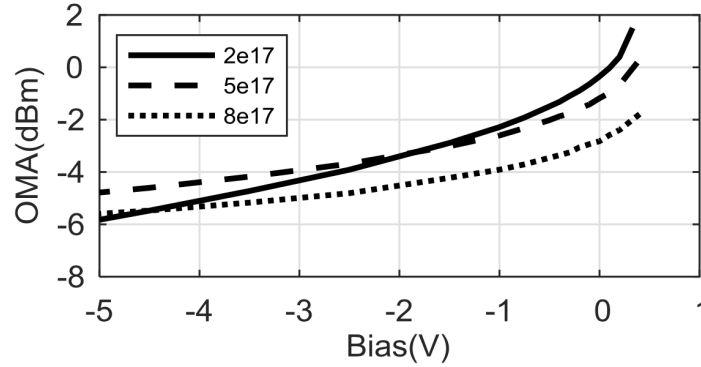


Figure 3.26: Optical modulation amplitude against bias voltage for various doping levels in a silicon MZM with 1.5 mm phase shifters. The applied peak-to-peak RF voltage is 1 V

decrease the achievable ER.

This trade-off can be analysed through the OMA produced by the device for a given V_{pp} . Figure 3.26 shows the OMA against bias point for the three doping levels when a V_{pp} of 1 V and an optical input power of 10 dBm are applied, assuming a combined 6 dB extra loss from input and output grating couplers. For all cases, the OMA is maximised at low reverse bias. Although free-carrier absorption is greater here, the increased phase shift per volt from the PN junctions (and hence, increased ER) outweighs the penalty of larger IL. In addition, we note from Figure 3.25 that the chosen V_{pp} is far below the half- V_{π} of the device, for all bias conditions. Despite this, the best OMA is achieved with the low doping case, provided the reverse bias voltage does not exceed -2 V. This shows that the low-doped PN junctions are favourable for SiP MZMs, even at the short lengths required for lumped modulators, as the IL rises too quickly with increased doping.

As the bitrate demanded from optical modulators rises, lumped MZMs face a challenge. The maximum length for which the device can be considered to be free from travelling-wave effects decreases as the frequency of the modulation (symbol rate) increases. Hence, multilevel modulation formats such as PAM4 and PAM8 are potentially favourable for lumped MZMs as they increase the bitrate without affecting the symbol rate. However, the use of multiple symbol levels presents another difficulty. As the phase shift produced by the PN junctions varies non-linearly with voltage, the SiP MZM will output data with mismatched symbol level spacing. A suitable metric to describe this is the level mismatch ratio (LMR), as defined by Equation 3.11. The LMR will be more impacted when the

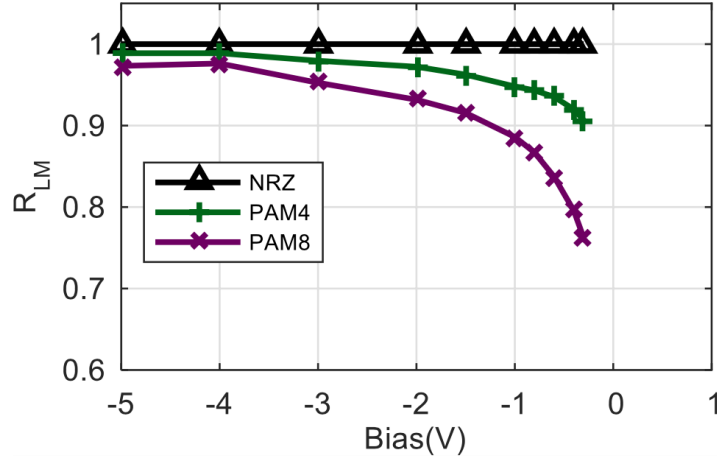


Figure 3.27: Level mismatch ratio in PAM4 and PAM8 signals, plotted against bias voltage for a silicon MZM with 2 mm phase shifters. The peak-to-peak RF voltage is 2 V. The NRZ signal is shown for purpose of comparison.

phase shift produced by the PN junctions is large. Therefore, we simulate here a device with a relatively long length of 2 mm and set the peak-to-peak voltage V_{pp} to 2 V. Figure 3.27 plots the resulting LMR against bias voltage for PAM4 and PAM8 modulation formats (with NRZ as a reference). We note that LMR is degraded more at lower reverse biases, where the phase shift varies the most from linearity. PAM4 is compromised less than PAM8, with their minimum LMR values reaching 0.9 and 0.76, respectively.

An impaired level separation mismatch ratio provides an additional penalty alongside the IL and ER penalties, but its effect on system performance is not captured by the OMA. Hence, to optimise the trade-off of IL, ER and LMR, we investigate their impact on the system by defining a transmission link penalty, in dB, as

$$TLP = IL + [ER \text{ penalty}] + [LMR \text{ penalty}]. \quad (3.21)$$

We then seek to minimise this as a figure of merit. This is similar to the approach taken in [130], except here we include the effect of mismatched symbol levels. The first penalty, IL, is easily obtained from Figure 3.25. The ER and LMR penalties are calculated numerically, as follows for a PAM format with M symbol levels. Assuming thermal noise dominates, the probability of receiving symbol j when symbol i is transmitted is

$$P_{ij} = \frac{1}{2} \operatorname{erfc} \left(\frac{I_{th,j} - I_i}{\sigma \sqrt{2}} \right) + \frac{1}{2} \operatorname{erfc} \left(\frac{I_i - I_{th,j+1}}{\sigma \sqrt{2}} \right), \quad (3.22)$$

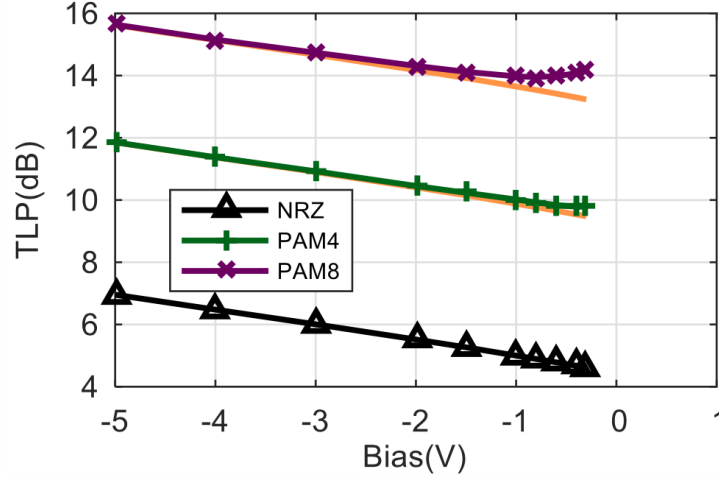


Figure 3.28: The transmission link penalty for NRZ, PAM4 and PAM8 signals, plotted against bias voltage for a silicon MZM with 2 mm phase shifters. The peak-to-peak RF voltage is 2 V. Orange lines represent the TLP when $R_{LM} = 1$.

where I_i is the receiver photocurrent at symbol i , $I_{th,j}$ the threshold current for symbol j , and σ is the root mean square value of the noise current. For a Gray-coded system, the resulting bit-error rate (BER) can be expressed as

$$BER = \frac{1}{M} \sum_{i=0}^{M-1} \sum_{j=0, j \neq i}^{M-1} \frac{1}{\log_2(M)} P_{ij}. \quad (3.23)$$

Taking the received symbols levels from the model, the average power into the receiver is swept until a desired reference BER (10^{-9}) is obtained. This is repeated for an ideal signal with infinite ER and perfect LMR, and the power penalty between the two signals is established.

Figure 3.28 plots this figure of merit, the TLP, against bias for NRZ, PAM4 and PAM8 formats. In all cases, the TLP reduces with smaller reverse bias, showing that the benefit of lowered half- V_π dominates over the increased insertion loss and degraded LMR. To display the impact of LMR on signal quality, the TLP is plotted both including and excluding the LMR penalty. Interestingly, even for PAM8, which is the most compromised format, LMR has only a 1 dB contribution to the TLP. This indicates that at low doping levels, PAM4 and PAM8 formats can be used for SiP MZMs without being significantly impacted by the effects of mismatched symbol levels.

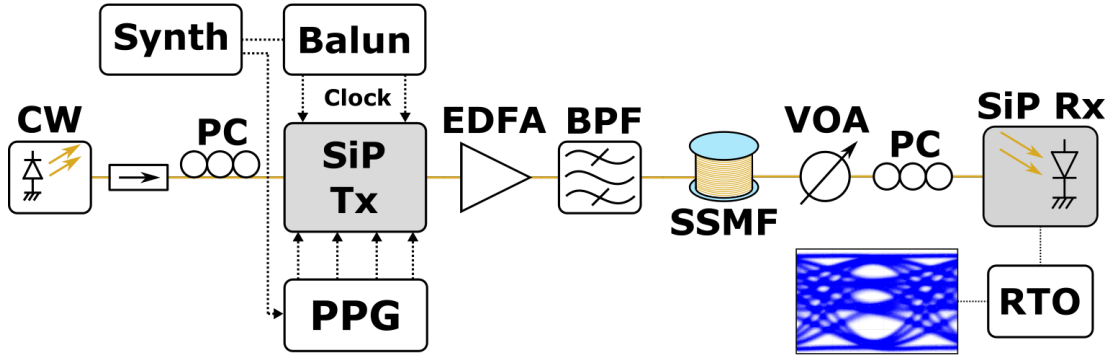


Figure 3.29: Full SiP link, including switched-capacitor PAM4 lumped MZM Tx, and silicon photoreceiver Rx.

3.5 Full Silicon Photonic Link Demonstration

The following sections focus on the experimental demonstration of a full silicon photonic transmission link involving a switched-capacitor, lumped MZM, which was published in [131]. This was the first demonstration of such a transmission scheme using both a silicon photonic MZM and a silicon photonic photoreceiver. Optimisation of the operating conditions are also shown through additional simulations.

3.5.1 Experimental Demonstration

The experimental setup is shown in Figure 3.29. Light from a continuous-wave (CW) C-band laser is passed through an optical isolator and polarisation controller (PC) before being injected into the switched-capacitor-driven silicon MZM (SiP Tx) via a grating coupler. The electronic driver circuit uses a differential RF clock and two sets of differential RF NRZ signals to produce a 2.4 V peak-to-peak differential PAM4 drive signal for the SiP MZM. The 15 GHz differential clock is provided by a frequency synthesiser (Synth) which has its output split by a balun. This clock signal simultaneously triggered a pulse pattern generator (PPG) which provided the two differential NRZ signals of PRBS7 and PRBS11, respectively. The differential clock and NRZ signals were aligned correctly in time for input to the SiP Tx package using discrete RF delay lines. The NRZ signals are combined in the switched-capacitor driver to form a PAM4 electrical signal that is then applied to the each arm of the MZM in dual-drive, push-pull configuration. The DC bias for each MZM arm was applied directly to the RF outputs of the driver via the on-chip resistors.

After undergoing modulation in the SiP Tx, the resulting optical 30 Gb/s PAM4 signal is coupled out of the chip via another grating coupler. Due to the large insertion loss of the silicon grating couplers used in this prototype device (approximately 5 dB each), an erbium-doped fibre amplifier (EDFA) is needed to boost the optical power of the signal. By compensating for any insertion loss (IL) in the MZM, the EDFA also provides an option for operating the MZM below its quadrature point. Doing so moves the operating point of the MZM towards the null of the transfer function, thereby increasing the extinction ratio (ER) of the signal but also resulting in greater IL and level mismatch ratio (LMR). With EDFAs typically capable of providing in excess of 30 dB gain, the increased IL can be cancelled by the EDFA resulting in a net gain in optical modulation amplitude (OMA), which, provided the LMR has not degraded too much, leads to greater receiver sensitivity [132]. In this experiment, the unbalanced MZM was operated below quadrature by tuning the laser wavelength to 1550.55 nm, resulting in a relatively large extinction ratio of 7 dB and an LMR of 0.78. However, it should be noted the use of optical amplification here was only for experimental purposes; it is anticipated that future iterations of the device with improved insertion losses will remove the need for optical amplification in the link. Therefore, the performance of the link without optical amplification is also evaluated in Section 3.5.2, following the discussion of these experimental results.

After being amplified, the optical signal is passed through a 0.8 nm bandpass filter (BPF) to remove out-of-band amplified spontaneous emission (ASE) noise from the EDFA. It is then transmitted through several lengths of standard single-mode fibre (SSMF) before a variable optical attenuator (VOA) is used to control the power at the input of the SiP Rx. The electrical PAM4 output of the SiP Rx was captured on a real-time oscilloscope (RTO), which had a sample rate of 80 GS/s and a bandwidth digitally limited to 15 GHz to remove higher-frequency noise. The SiP Rx is a linear photoreceiver that was designed and fabricated as part of another PhD thesis in our research group. Full details of the device can be seen in [133]. Essentially, it consists of a high-speed silicon-germanium photodiode with a 3 dB electro-optic bandwidth of approximately 30 GHz, and an electronic circuit comprised of multiple stages, including a transimpedance amplifier (TIA), a variable-gain amplifier (VGA), a continuous-time linear equaliser (CTLE), and a 50 Ω output buffer for termination. The CTLE was not used in this investigation. The electrical circuit in the photoreceiver has a 3 dB bandwidth of approximately 14 GHz.

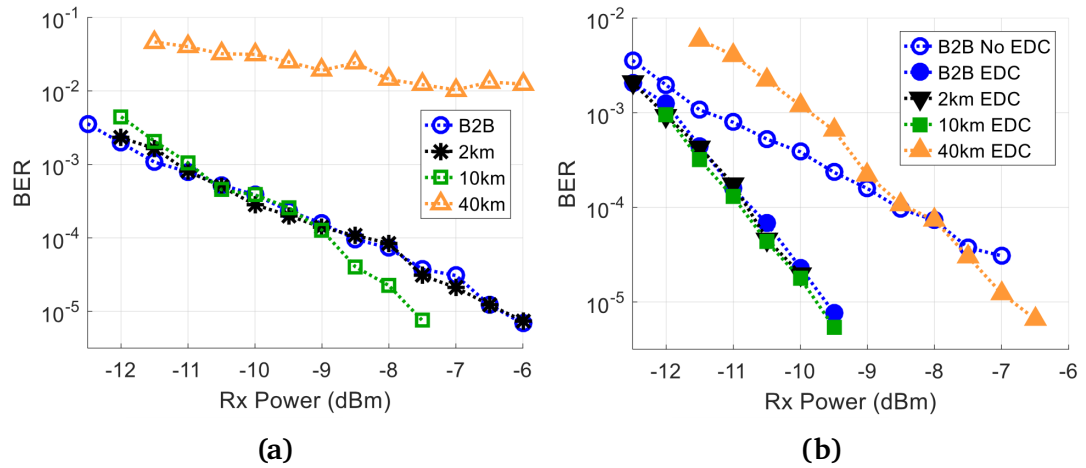


Figure 3.30: Measured BER curves for SiP PAM4 link, for various lengths of SSMF, with (a) no equalisation applied and (b) 14-tap fractionally spaced FFE applied.

The performance of the optical link was investigated by sweeping the power and measuring the corresponding bit error rate (BER) for various lengths of SSMF: back to back (B2B), 2 km, 10 km and 40 km. The power at the SiP Rx was swept from -13 dBm to -6 dBm in steps of 0.5 dB using the VOA. The resulting signal at the RTO was saved for offline analysis using Matlab, where eye diagrams were produced and the BER obtained via error counting. A plot of the BER is shown in Figure 3.30(a) for the various SSMF lengths. Taking the FEC limit as 10^{-3} , a small (<0.5 dB) power penalty was observed between the B2B, 2 km and 10 km cases, demonstrating the feasibility of the SiP transmitter and receiver for intra-data-centre applications. We note that towards higher received powers (>-9 dBm), the 10 km link has a better measured BER than the B2B and 2 km links. This is likely due to slight chirp in the silicon MZM being compensated by the positive dispersion coefficient of the fibre. The fact that this performance improvement does not hold for optical powers lower than -9 dBm is probably due to slight shift in the operating conditions of the Tx or Rx over the course of this BER sweep, such as a temperature or polarisation change. For the 40 km link, there is a large degradation in performance due to the significant chromatic dispersion present over the fibre, with the BER clearly flooring above the FEC limit. This indicates that the SiP link, in this form, is not suitable for *inter*-data-centre communications.

However, electronic dispersion compensation (EDC) can be employed using digital equalisation at the receiver in order to extend the reach. To demonstrate this, a fractionally spaced, feed-forward equaliser (FFE) with 14 taps is applied

digitally to the captured waveforms. Essentially, the FFE creates a number of delayed versions of the waveform equal to the number of taps, and adds them back to the original waveform after scaling each by its tap weight. The tap weights are adjusted by a gradient descent algorithm to minimise intersymbol interference.

This FFE increases the operational optical link distance as shown in Figure 3.30(b). The B2B, 2 km and 10 km curves are almost identical to each other after equalisation, and have improved performance compared to the non-equalised B2B case. As these link lengths were not seriously compromised by dispersion before equalisation, this improvement can be attributed to the FFE compensating for bandwidth impairments in the electronic and optical components of the Tx or Rx. There is a marked improvement in the performance of the 40 km link when the FFE is applied, with no error floor exhibited in the range of received powers and measured BER reaching 10^{-5} , indicating that the SiP Tx and Rx used are suitable for inter-data-centre communications if digital equalisation is applied.

The SiP Tx consumes 216.7 mW from 1.2 V and 2.4 V supplies, representing a power efficiency of 7.2 pJ/bit. The SiP Rx consumes 85.8 mW from two 1.2 V supplies, giving a power efficiency of 2.9 pJ/bit. Combined, without considering the equalisation, which was investigated in post-processing as a hypothetical scenario, the two devices consume 10.1 pJ/bit when operating with a 30 Gb/s PAM4 signal. The measured energy consumed per bit is noticeably higher than the 1 pJ/bit touted as a goal for data centre links back in Chapter 1. However, significant savings can be made in the power consumption of the switched capacitor driver, as discussed by the circuit designer in [128]. Most notably, the voltage doublers and level shifters in the circuit, which consume over 80 % of its power, could be made more efficient through future designs, the details of which are beyond the scope of this thesis.

3.5.2 Operation of the SiP Link without Optical Amplification

As discussed in the previous section, the SiP MZM was operated below its quadrature point to achieve an increase in ER. However, given that optical amplification (particularly EDFAs) are power hungry components, their use in data centre settings is not guaranteed. Therefore, this section analyses through simulations the performance of the SiP link in the absence of any optical amplification. It was not possible to carry out this investigation experimentally as

the prototype switched-capacitor MZM suffered an unrepairable fibre coupling failure subsequent to the measurements carried out in Section 3.5.1. The simulated MZM here uses phase shifters of length 1.5 mm and is set to have equal doping levels of $5 \times 10^{17} \text{ cm}^{-3}$ as a typical, representative example of a moderately doped MZM. The peak-to-peak RF drive voltage across the phase shifters is taken as 1.2 V.

Firstly, it should be established that, without optical amplification, the optimum operation point on the MZM's transfer function is quadrature. As the peak-to-peak swing of the electrical signal generated at the photoreceiver is proportional to the OMA of the optical signal, it will here suffice to maximise the OMA rather than undertaking the more computationally intensive process of sweeping the BER for all operating points. The three main methods of varying the operating point on the MZM's transfer function are tuning the input optical wavelength (in the case of an unbalanced MZM), applying a thermal phase shift in one arm, and providing unequal bias voltages to each arm of the MZM.

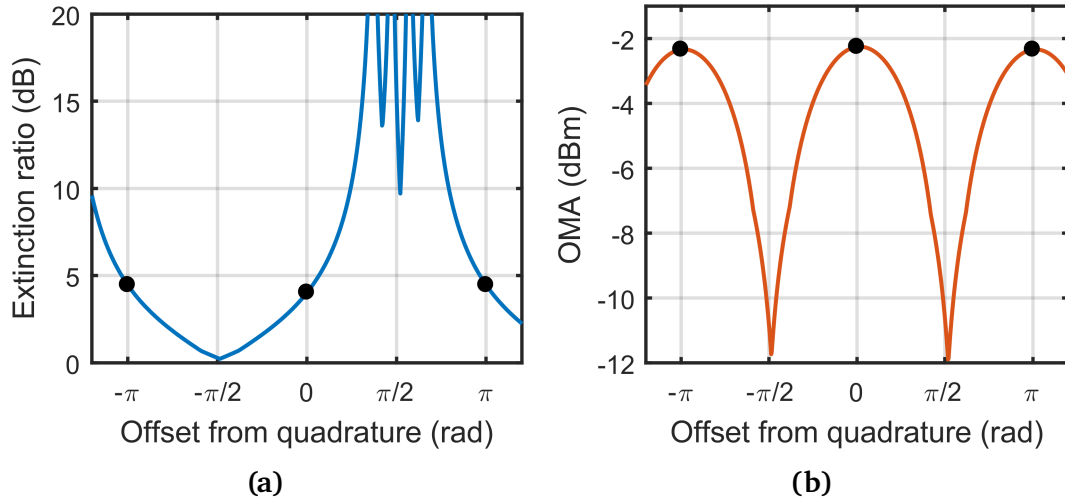


Figure 3.31: (a) ER and (b) OMA against thermal or wavelength-tuned phase offset from quadrature in the MZM.

Using thermal or wavelength tuning are essentially equivalent, as they both give direct control of the operation point of the MZM through phase without affecting the carrier depletion characteristics of the PN junctions. The effects of tuning the MZM's operation point in this fashion on ER and OMA are shown in Figures 3.31(a) and 3.31(b), respectively. If the operation point is offset from quadrature towards the peak of the MZM transfer function, the ER drops, as the ratio between rails of the signal decreases. Contrastingly, if we move towards the null of the transfer function, the ER rises. (The sharp spikes in ER around $\pi/2$

rad occur due to the PAM4 symbol levels transitioning over the transfer function null.) Significantly, however, the optimum operation point in terms of OMA is quadrature, as any increase in ER towards the null is counteracted by increased insertion loss, and vice versa for a move towards the peak.

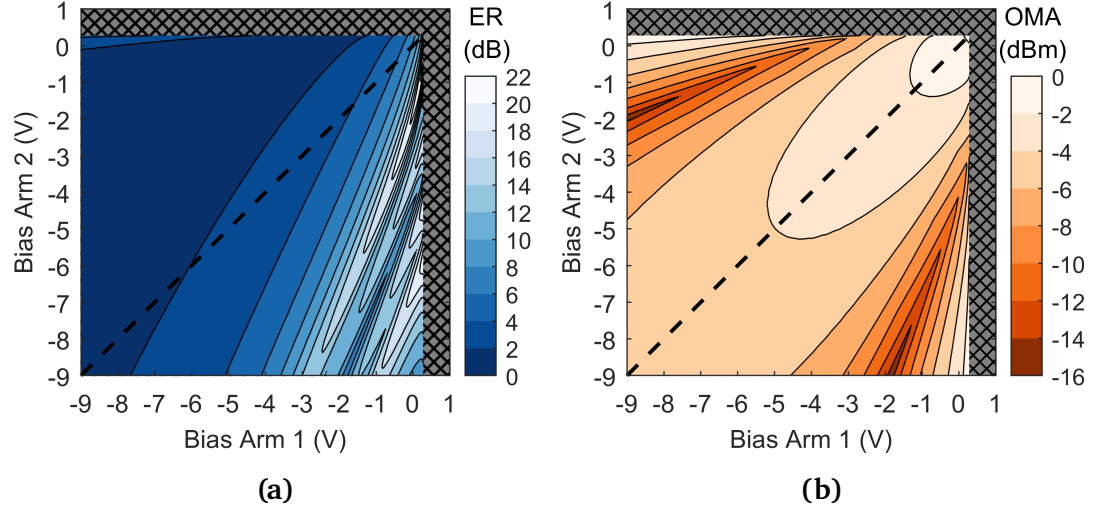


Figure 3.32: (a) ER and (b) OMA against bias voltage in the two MZM arms. The dotted lines mark equal bias voltages and correspond to quadrature operation.

Tuning the operation point through the bias voltages of the PN junction phase shifters is slightly different to thermal or wavelength tuning methods, as the depletion-mode phase shifters have different phase and absorption characteristics depending on the applied voltage. Therefore the various combinations of bias voltages in the two arms should be swept and examined. The resulting ER and OMA are shown in Figures 3.32(a) and 3.32(b), respectively, where the MZM is initially laced at quadrature through thermal or wavelength tuning under 0 V bias in each arm, and the bias voltages are then swept. The grey, hatched regions indicate bias points that are not of interest as the peaks of the RF voltage swing would exceed the built-in voltage V_b of the PN junction and thus require modelling of the junction under a different regime, i.e., forward bias. The black, dotted line marks the points of quadrature operation, when the bias voltages of the two arms are equal. Similarly to Figure 3.31(a), where the operating point was tuned by means of a thermal or wavelength change, in Figure 3.32(a) there are regions either side of quadrature where the ER increases (due to moving towards the transfer function null) and decreases (due to moving towards the transfer function peak). Figure 3.32(b) is clear, however. Despite the slightly more nuanced behaviour of SiP MZM phase control with bias voltage, the same result as for thermal and wavelength tuning holds: any move away from quadrature

results in reduced OMA. Furthermore, similarly to the results of Figure 3.26, the optimum operating conditions for maximum OMA are quadrature at low reverse biases, where the PN junctions provide the greatest phase shift.

With quadrature established as the optimum operation point in this context, the performance of the full 15 Gbaud PAM4 SiP link over fibre is then re-evaluated through a Matlab simulation testbed that was developed in-house. The phase shifters are modelled with the same parameters as used for Figures 3.31 and 3.32, each under reverse bias of -1 V, with peak-to-peak RF voltages of 1.2 V. The resulting optical PAM4 signal from the MZM has an ER of 4.8 dB. The frequency response of the switched-capacitor MZM is modelled simply as a fourth-order Bessel response with a 3 dB bandwidth equal to the symbol rate. The receiver is modelled with its specified responsivity of 0.8 , an NEP of 29.25 pW/ $\sqrt{\text{Hz}}$ and a fourth-order Bessel response that follows its measured 3 dB bandwidth of 14 GHz. The modelled fibre has a dispersion parameter of 17 ps/nmkm in C-band.

The input power at the receiver is swept for lengths of fibre from 0 km to 40 km and the BER is obtained in each case through error counting. The resulting receiver sensitivity at FEC threshold (taken as a BER of 10^{-3}) is shown against fibre length in Figure 3.33. As the ER of the simulated quadrature case is 4.8 dB compared to the 7 dB ER of the measured below-quadrature case, the receiver sensitivity could therefore be expected to be worse for the simulation. The typical expression for the sensitivity penalty in a PAM signal due to finite extinction ratio is given in dB by [39]

$$\delta_{ER} = 10 \log_{10} \left(\frac{ER_{lin} + 1}{ER_{lin} - 1} \right), \quad (3.24)$$

where ER_{lin} is the extinction ratio expressed in linear terms rather than dB. Filling in the extinction ratios of 4.8 dB and 7 dB for the two cases to this expression, the expected sensitivity penalty for the reduced extinction ratio at quadrature is calculated to be 1.2 dB.

However, the measurements in back-to-back configuration are similar for both, with the simulated quadrature case actually slightly outperforming the experimental below-quadrature case (sensitivities of -12 dBm and -11.5 dBm, respectively). This can be attributed to the degraded LMR of the signal in the below-quadrature case, along with the impact of magnitude and phase variations in the actual device frequency responses not captured by the ideal Bessel responses used in the simulation. Nevertheless, the simulation does

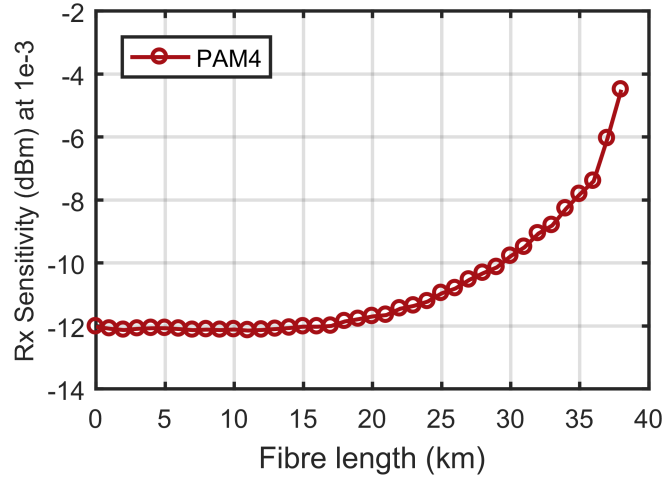


Figure 3.33: Rx sensitivity for the simulated 15 Gbaud SiP switched-capacitor PAM4 link at quadrature.

indicate that a well-designed switched-capacitor SiP link would operate successfully at quadrature without optical amplification. Like the experimental results, the performance of the link over fibre is not sufficient to reach 40 km transmission without implementing additional receiver-side equalisation. It does, however, reach quite close to this distance, indicating that the switched capacitor design could be reasonably feasible for intra-data-centre communications if greater RF voltage swing could be made available from the driver, or if more advanced silicon phase shifter designs with lower V_π were used.

3.6 Summary

This chapter focusses on the characterisation and optimisation of lumped silicon photonic MZMs, both in terms of doping and drive conditions. Section 3.2 begins with a description and measurement of the VI characteristics of the PN junction phase shifters in a lumped MZM. A method is then described for measuring the voltage dependent phase shift and absorption produced by these PN junctions using the unbalanced design of the MZM arms. Flat-spectrum SPHS are shown to be extremely convenient for measuring the frequency response of the MZM without needing to use a VNA. This is particularly useful for later experiment setups in Chapter 4.

Section 3.3 introduces a mixed numerical-analytical modelling approach that is used to determine the dependence of phase and absorption characteristics of the PN-junction phase shifters with doping and bias levels. This model will be

used throughout the thesis for various device and system-level simulations. In this section, the model is first used to predict the insertion loss, ER, and LMR (for PAM4) of lumped SiP MZMs in a general sense, for various drive voltage conditions. A simple analytical model is also described for the RC bandwidth of lumped MZMs, and various doping conditions and source impedances are analysed to identify key trends in bandwidth. Crucially, drivers with $50\ \Omega$ source impedance are shown to dramatically hamper the RC bandwidth of lumped MZMs. This leads to the investigation of low-impedance drivers to boost the bandwidth of lumped MZMs.

Section 3.4 focusses on the optimisation of a novel low-impedance, switched-capacitor driver for lumped MZMs. The switched-capacitor driver allows convenient generation of PAM4 signals without the use of a traditional electrical DAC, and is designed to significantly increase the bandwidth of a lumped MZM. As SiP phase shifters have voltage-dependent, non-linear phase shift and absorption, the effect of their doping and bias levels on the performance of the switched-capacitor lumped MZM should be analysed. This done in this section through the phase shifter model described earlier in this chapter. Various doping levels for the phase shifters are compared, and a relatively low doping of $2 \times 10^{17}\ \text{cm}^{-3}$ is shown to be optimal. The LMR introduced due to the SiP phase shifters in this switched-capacitor scheme is shown to be negligible.

Section 3.5 then demonstrates the operation of a fully packaged lumped MZM with switched-capacitor driver for the first time in a silicon photonic link. Successful transmission below the FEC limit is shown for 30 Gb/s operation up to 10 km, and receiver-side equalisation is shown to enable further transmission at 40 km. The ER in this setup was increased by biasing the MZM below quadrature and post-amplifying with an EDFA. For a more accurate representation of the link in a data centre setting, without optical amplification, simulations are used to show that the optimal operation point on the MZM's transfer function is quadrature, and that the switched-capacitor MZM would still successfully operate for data centre links.

Chapter 4

Predistortion for High-Speed Lumped SiP MZMs

4.1 Introduction

This chapter investigates the potential for using electrical predistortion or precompensation in driver circuits to facilitate high-speed modulation of lumped SiP MZMs. Firstly, in Section 4.2, an experimental demonstration of 50 Gb/s PAM4 and EDB signals is given using precompensated lumped SiP MZMs. As the precompensation used there was relatively simple, additional receiver-side equalisation is needed to reach 50 Gb/s transmission below FEC threshold. Section 4.3 shows how an optimal, matched precompensation filter is obtained, and also provides a simple method for evaluating the increased power dissipation as a result of the larger voltage required for precompensation. Section 4.4 then shows through experiments and simulations how optimally designed precompensation removes the need for the receiver-side equalisation to obtain 50 Gb/s performance with the lumped MZM.

As the measured bandwidth of the lumped MZM available for these experiments is quite low, in Section 4.5 an equivalent circuit model for the device is determined and fitted to S11 measurements to facilitate modelling of a device with better performance. Attention also is paid to the effect of the optical transit time of the light through the MZM on the device's frequency response. Using the equivalent circuit model, the system performance, increased drive voltage and energy dissipation, when precompensation is applied to the lumped MZM, are evaluated. These trends in performance are also investigated for lumped MZMs

with low-impedance drivers, with and without precompensation. Consideration is also given to the onset of significant inductive peaking in the lumped MZM as the source impedance is lowered. Finally, the potential advantages of using precompensation with a low-impedance driver for lumped MZMs is identified.

4.2 Experimental Evaluation of Precompensation for Lumped MZMs

The work in this section is an extension of our work which was published and presented in [134]. That work evaluated the performance of a SiP TW-MZM with and without its $50\ \Omega$ termination, where precompensation was used to boost the bandwidth when the electrodes were left open. The unterminated TW-MZM was used as a representative example of a lumped MZM to allow direct comparison of performance with and without the termination. In the extended work presented in this section, a true lumped-design MZM (of the same length) is added to the comparison, to evaluate if there is a significant difference between simply using an unterminated TW-MZM as a lumped device, as opposed to a more standard lumped MZM.

4.2.1 Travelling-Wave and Lumped MZM Frequency Response Comparison

Microscope photographs of the TW-MZM and the lumped MZM used in these experiments were previously shown in Figure 2.14. The devices were noted to have clear differences in their electrode design; the TW-MZM uses co-planar waveguides to minimise the frequency-dependent transmission line effects of RF attenuation, impedance mismatch and velocity mismatch between the RF and optical waves, while the lumped MZM uses thinner, simpler electrodes.

In following work, we compare the performance of the TW-MZM with a $50\ \Omega$ termination against two different lumped MZMs. The first one is obtained by not simply terminating the TW-MZM of Figure 2.14(b), leaving the co-planar waveguide electrodes open at the far end of the electrodes. The second one is the typical lumped design in Figure 2.14(a). Both devices are 1.5 mm long.

Figure 4.1 shows the measured frequency responses of the TW-MZM with and without the $50\ \Omega$ termination (which we will refer to as terminated and unterminated, respectively), and the more typical lumped MZM (referred

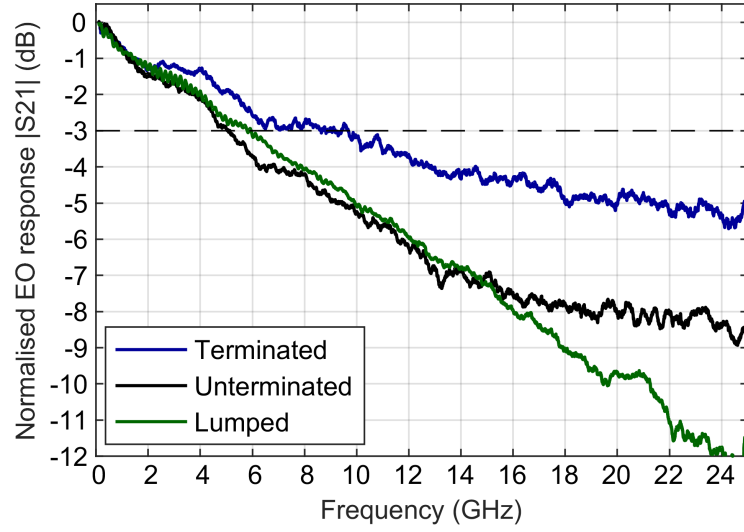


Figure 4.1: Normalised electro-optic frequency response of the terminated TW-MZM, unterminated TW-MZM, and lumped MZM.

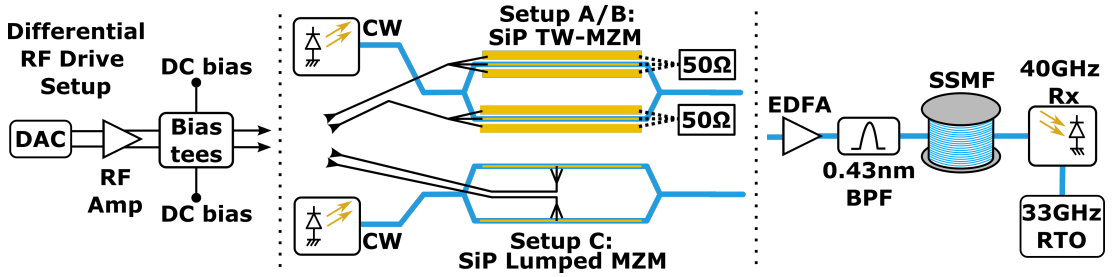


Figure 4.2: Experimental setup for 25 Gb/s and 50 Gb/s modulation with a terminated (setup A, 50 Ω load attached) and unterminated (setup B, 50 Ω load removed) SiP TW-MZM, and for a lumped SiP MZM (setup C).

to simply as the lumped MZM), each normalised to their lowest measured frequencies. As expected, the TW-MZM has the best frequency response, with a 3 dB bandwidth of 9.5 GHz. Removing the termination from the TW-MZM introduces RF reflections from the open-ended electrodes which severely impair the response, reducing the 3 dB bandwidth to 5 GHz. The lumped MZM initially has a slightly better frequency performance than the unterminated TW-MZM, with a 3 dB bandwidth of 6 GHz, possibly due to the electrodes being electrically contacted in the middle rather than the end. Noticeably, however, the lumped MZM's frequency response rolls off more steeply than the unterminated TW-MZM after 15 GHz. This can be attributed to greater RF attenuation at higher frequencies due to the lumped MZM's narrower electrodes.

4.2.2 Experiment Setup

The experimental setups used to characterise the 25 Gb/s and 50 Gb/s performance of the terminated and unterminated TW-MZM and the lumped MZM are depicted in Figure 4.2. The SiP MZMs have a V_π of 8 V to 10 V depending on the applied reverse bias. Differential NRZ, PAM4 and EDB electrical drive signals are generated by a differential 33 GHz, 100 GS/s AWG and DAC, and amplified by a commercial 32 GHz linear RF amplifier before being applied to the MZMs in dual-drive, push-pull mode via a GSGGSG probe. Identical DC bias was added to each RF drive signal via an external bias tee. The external 50 Ω terminating resistor is added and removed from the TW-MZM via another high-speed probe to create the terminated case (setup A) and unterminated case (setup B). No termination resistor is used for the lumped MZM (setup C).

A deliberate, small asymmetry in the MZM arm lengths allows quadrature operation to be reached by finely tuning the commercial C-band laser in a 2 nm range around 1550 nm, avoiding the need for DC or thermal bias sections in the arms. This wavelength was optimised for each modulation format and bit rate due to different local heating of the phase shifters changing the exact quadrature point. Due to the high loss of the grating couplers (approximately 6 dB each) on the devices, the output of the MZM is boosted by an erbium-doped fibre amplifier (EDFA) for our analysis. However, CMOS-compatible gratings with much lower losses of around 1 dB have been demonstrated in the literature through the use of apodised (non-uniform) grating period and fill-factor along with optimisation of the grating coupler etching depth [135]. Grating couplers with such low insertion loss would remove the need for optical amplification. The optical signal is filtered using a 0.43 nm bandpass filter (BPF) to suppress the out-of-band EDFA amplified spontaneous emission before being transmitted in both back to back (B2B) and over 10 km of SSMF. A variable optical attenuator (VOA) is used to sweep the received optical power into a commercial 40 GHz photoreceiver. The resulting electrical signal is sampled by a 33 GHz, 80 GS/s real-time oscilloscope (RTO) and analysed offline using Matlab to measure bit error rate (BER) and generate eye diagrams. During the offline processing, the Rx bandwidth is restricted to 0.75 times the symbol rate for the NRZ and PAM4 cases, respectively, and 0.4 times the symbol rate for the EDB case. These values were selected after sweeping the receiver bandwidth and analysing the resulting BER to determine the best balance of thermal noise and inter-symbol

interference for the modulation formats.

To obtain the frequency response of the system, flat-spectrum SPHS are generated by the AWG and sent through the link. By separately measuring the response of the DAC, RF amplifiers and cables, the frequency responses of the MZMs are extracted. The effect of Tx-side DSP (predistortion) on the frequency response is analysed by applying it as a digital filter to the SPHS before transmission.

4.2.3 Results

The low bandwidth exhibited by the unterminated and lumped MZMs in Figure 4.1 means that 25 Gb/s NRZ modulation should be unachievable for such devices. However, DSP can compensate for a non-ideal frequency response through predistortion of the signal at the transmitter (Tx) and/or equalisation at the receiver (Rx). To illustrate this, Figures 4.3(a) and 4.3(b) show the frequency response of the unterminated and lumped MZMs when a simple first-order finite impulse response (FIR) filter (two taps, weighting: $[1 \ -0.25]$), symbol-spaced at 25 Gb/s, is applied by the DAC. The 3 dB bandwidth is in each case now approximately 12 GHz, closely matching the terminated MZM.

Figure 4.4 compares the back-to-back BER performance of the terminated, unterminated and lumped MZM for 25 Gb/s NRZ, where we have applied the FIR at the Tx for the unterminated and lumped MZMs. No predistortion is applied to the terminated MZM. For a fair comparison, both signals have the same extinction ratio of approximately 6 dB. It is important to note that the drive voltage applied per MZM arm to reach this ER was considerably lower for the unterminated and lumped cases (approximately 3 V peak to peak after the RF amplifiers) than the terminated case (approximately 6 V peak to peak). The eye diagrams shown in Figure 4.5(a) for the terminated MZM, Figure 4.5(b) for the FIR-filtered unterminated MZM and Figure 4.5(c) for the FIR-filtered lumped MZM are relatively similar. Each exhibits comparable eye closure, implying that a similar effective distortion due to bandwidth restriction is present in each case. They also give similar BER performance as shown in Figure 4.4. The terminated and lumped MZMs perform equally well, with both reaching the HD-FEC limit at an Rx power of -14 dBm and the KP4-FEC limit at -12.5 dBm, while the unterminated MZM shows a small sensitivity penalty of less than 1 dB.

In addition to NRZ modulation, we also investigated the performance of the unterminated and lumped MZMs with Tx-side DSP for 25 Gb/s PAM4 and EDB.

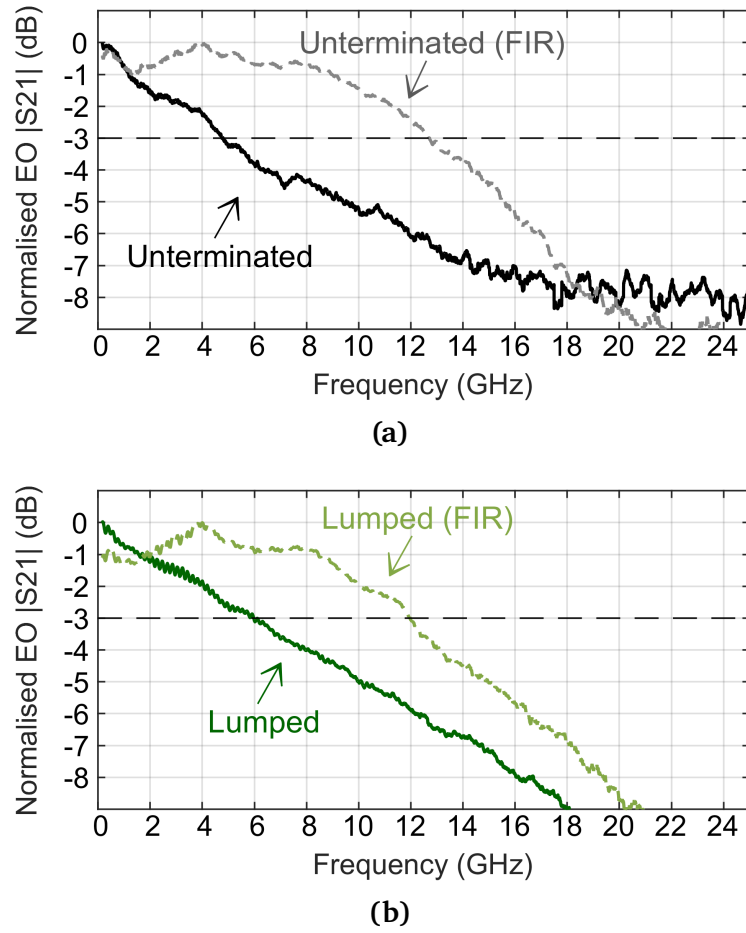


Figure 4.3: Effect of the FIR predistortion filter on (a) the unterminated MZM and (b) the lumped MZM.

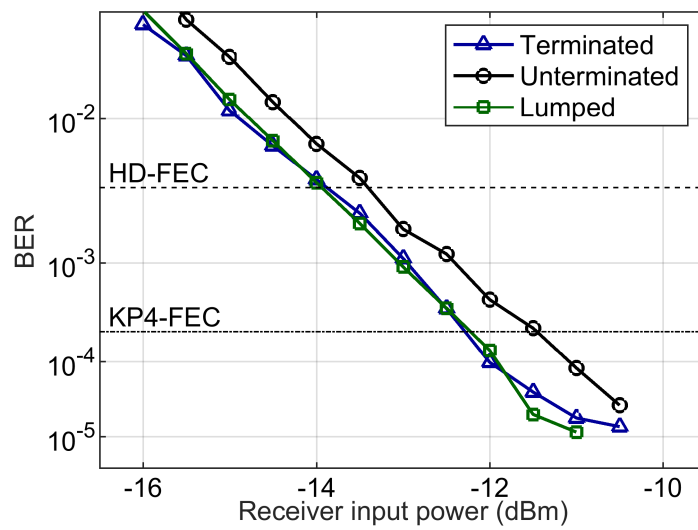


Figure 4.4: BER comparison of the unterminated MZM with no predistortion applied, terminated MZM with the first-order FIR predistortion applied, and lumped MZM with same predistortion applied for 25 Gb/s NRZ modulation.

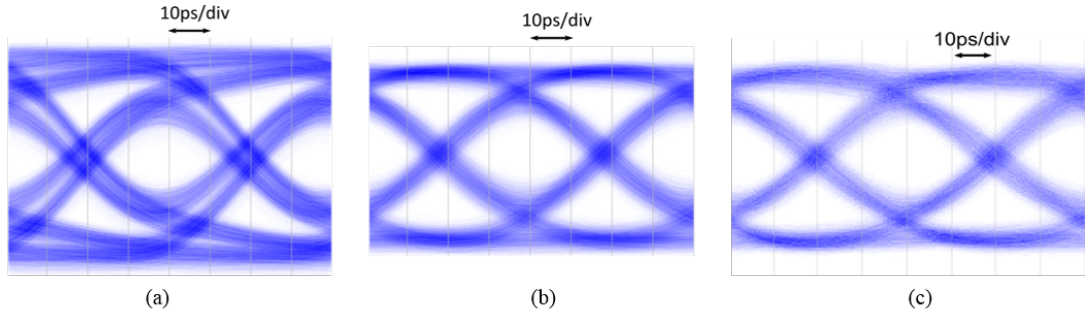


Figure 4.5: (a) 25 Gb/s terminated MZM NRZ eye. (b) 25 Gb/s unterminated MZM NRZ eye with first-order FIR Tx filter. (c) 25 Gb/s lumped MZM NRZ eye with first-order FIR Tx filter.

The BER results for the unterminated MZM are shown in Figure 4.6(a), with its 25 Gb/s NRZ BER curve shown as a reference. The PAM4 and EDB eye diagrams are shown for the unterminated MZM in Figure 4.6(b) and Figure 4.6(c), respectively. Each modulation format has an ER of approximately 6 dB. The same first-order FIR filter used for NRZ is applied to the 25 Gb/s EDB. Of the three modulation formats, EDB shows the best performance, with a 1 dB and 2 dB improvement over NRZ at the HD-FEC and KP4-FEC limits, respectively. For 25 Gb/s PAM4, a simple first-order FIR cannot adequately compensate the frequency response of the unterminated and lumped MZMs to clearly distinguish the four symbol levels over two. (Essentially, the smaller, multiple eye openings in the PAM4 signal are less tolerant of crossing-point impairments). To fully compensate the signal, a filter with a sample rate of twice the symbol rate is required to affect frequencies up to the Nyquist rate of the signal. Hence, we use a full precompensation technique by inverting the MZM's frequency responses from Figure 4.1 and applying them appropriately as a digital filter to the DACs. These resulting zero-forcing filters act to flatten the overall frequency of the link, essentially boosting high frequencies and attenuating low frequencies. Even after full precompensation, PAM4 displays a 5 dB and 7 dB penalty at the HD-FEC and KP4-FEC limits, respectively, in comparison to the first-order FIR NRZ case.

As shown in Figure 4.6, the lumped MZM exhibits similar performance to the unterminated MZM for 25 Gb/s PAM4 and EDB, with their corresponding BER curves matching to within 1 dB. For EDB, the lumped MZM reaches the HD-FEC and KP4-FEC limits at -14 dBm and -12 dBm, respectively; for PAM4, it reaches these limits at -8 dBm and -11 dBm, respectively. The BER results for the lumped MZM and the unterminated MZM differ with regards to whether NRZ outperforms EDB, but this difference is only 1 dB of sensitivity, reasonably

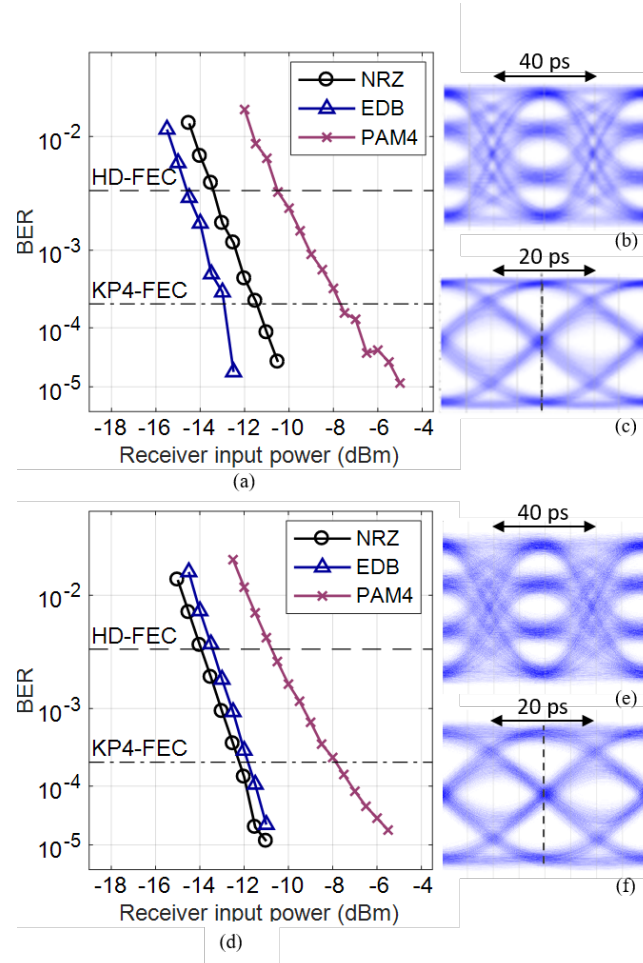


Figure 4.6: (a) BER performance of the unterminated MZM for 25 Gb/s NRZ, EDB and PAM4. (b) 25 Gb/s PAM4 eye diagram. (c) 25 Gb/s EDB eye diagram. The corresponding BER curves, PAM4 and EDB eye diagrams for the lumped MZM are shown in (d), (e) and (f) respectively.

within the bounds of error for measurements on separate devices. The PAM4 and EDB diagrams are shown for the lumped MZM in Figure 4.6(e) and Figure 4.6(f), respectively. The similarity in performance of the two devices is to be expected for 25 Gb/s modulation due to the close matching of their frequency response up to 15 GHz.

As mentioned in Section 4.2.1, for higher frequencies, the two devices start to show a marked difference in frequency response, with the lumped MZM presenting a steeper frequency roll-off compared to the unterminated TW-MZM. With this in mind, we investigate both modulators with a 50 Gb/s data rate in B2B and over 10 km of fibre to demonstrate feasibility of the system for the LR8 400G standard, which requires a power budget of 6.3 dB. At this data rate, the performance of NRZ modulation is severely limited by the bandwidth of

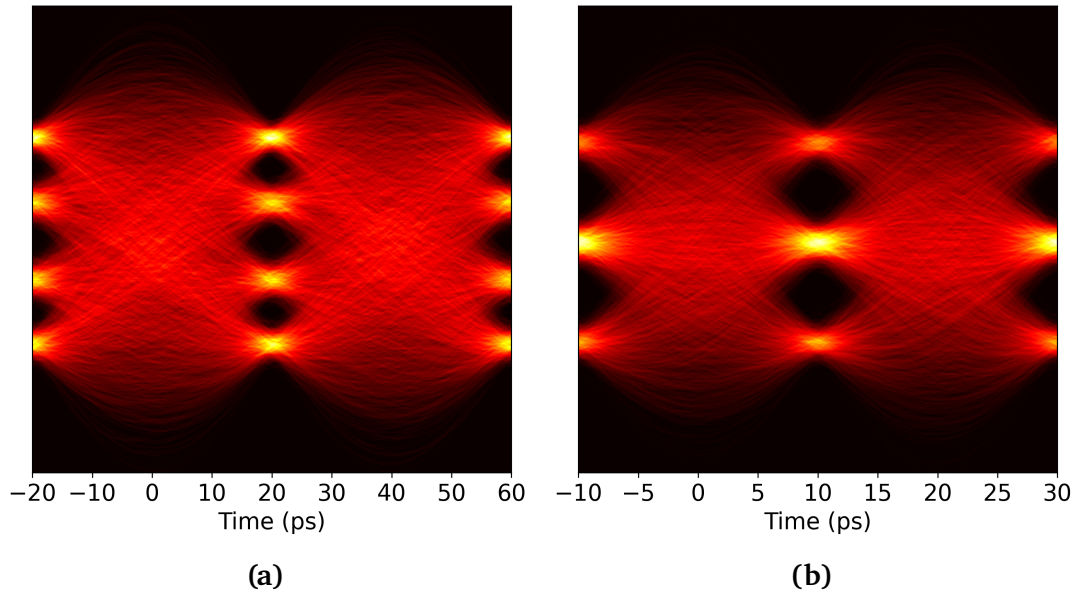


Figure 4.7: Eye diagrams obtained for the B2B 50Gb/s PAM4 and EDB signals using a 16-tap fractionally spaced FFE at the receiver.

the unterminated and lumped MZMs, and hence we focus on PAM4 and EDB modulation formats due to their higher spectral efficiency. EDB employs the same first-order FIR filter as for the 25 Gb/s case (now symbol-spaced at 50 Gb/s), while PAM4 again used the full frequency compensation. Rx equalisation is also applied for both formats in addition to the Tx predistortion in order to remove residual intersymbol interference from the bandwidth restriction and compensate for the dispersion introduced by the fibre propagation. The Rx equaliser used is feed-forward fractionally spaced over 8 symbols (FFE $8T/2$; 16 taps).

For continuity, Figures 4.7(a) and 4.7(b) show eye diagrams for the B2B 50Gb/s PAM4 and EDB signals with the fractionally spaced equalisation. The almost diamond-shaped eye openings are typical of an eye diagram obtained from this kind of equalised signal (where the receiver-side FFE algorithm produces one sample per symbol from the received waveform). They result as the signal is upsampled from one sample per symbol (below the Nyquist rate of the signal) to a higher sample rate to produce the eye diagram (using 400 samples per symbol in this case). The eye openings indicate relatively good signal quality, although this can be better analysed in the corresponding BER curves which follow.

Figure 4.8 shows the BER performance of the unterminated MZM for B2B and 10 km of SSMF. It should be noted that here, due to the high bandwidth roll-off of the unterminated MZM, the ER of the PAM4 signal drops to approximately

4.5 dB. Despite this, assuming a Tx optical launch power of 3 dBm, PAM4 comfortably achieves the LR8 power budget at the HD-FEC and KP4-FEC limits, with 3 dB and 5 dB margin, respectively. There is no significant power penalty between the B2B and 10 km PAM4 link, EDB outperforms PAM4 for 50 Gb/s 10 km transmission, with a 1 dB improvement in power budget at the FEC limits. However, fractionally spaced ($T/2$) equalisers require Rx circuitry operating at twice the clock rate. Such equalisation for 50 Gb/s EDB would require Rx circuitry operating at 100 GHz. To simplify such circuits, we also investigated symbol-spaced equalisation for EDB (FFE 8T; 8 taps). With this simpler equalisation, EDB shows a 1 dB power penalty from the fractionally spaced equalisation case, similar to the PAM4 performance, and therefore with a power budget 3 dB greater than the LR8 requirement.

Before comparing with the corresponding 50 Gb/s results for the lumped MZM, it should be recalled from Figure 4.1 that the lumped MZM has a more severe bandwidth roll-off than the unterminated MZM at frequencies higher than 15 GHz. The frequency precompensation used for the PAM4 signal is effectively boosting the high frequency content of the signal to try and compensate for the bandwidth restriction, but at the expense of the effective voltage swing of the driving signal at the output of the DAC. Because of the practical limitation on the gain of the driver amplifiers, this causes the ER of the precompensated 50 Gb/s PAM4 signal for the lumped MZM to drop further to 3.5 dB. The effect of this is noticeable when viewing the lumped MZM 50 Gb/s results in Figure 4.8(b). The PAM4 results for B2B clearly show the appearance of an error floor below the KP4-FEC limit while at 10 km, the error floor is above the KP4-FEC limit. The EDB curves reach the HD-FEC and KP4-FEC thresholds at Rx optical powers within 1 dB of the corresponding curves for the unterminated MZM.

The results for 50 Gb/s PAM4 show that the unterminated TW-MZM presents an advantage compared to the lumped MZM due to the better high frequency design of the electrodes. The improved performance at high frequency allows for a more efficient use of the transmitter precompensation without affecting the drive voltage swing as severely and, hence, the ER. On the other hand, the 50 Gb/s EDB shows very similar performance for the two modulators, although a suggestion of an error flooring effect is visible at higher powers. We attribute this to the fact that EDB is more robust to bandwidth restriction and can hence use the simpler FIR precompensation, which does not affect as severely the swing of the driving signals.

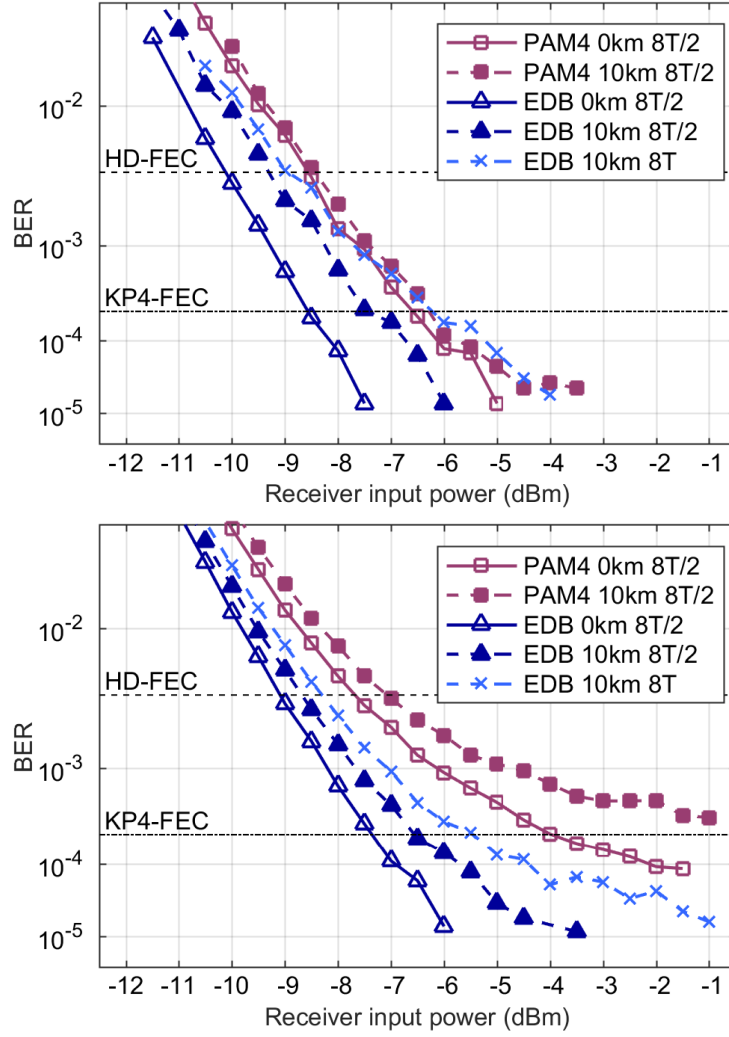


Figure 4.8: (a) The unterminated MZM's BER performance for 50 Gb/s EDB and PAM4 with fractionally spaced FFE over B2B and 10 km of fibre. Also shown is symbol spaced FFE for EDB over 10 km of fibre. (b) The corresponding curves for the lumped MZM, showing error flooring.

4.3 Designing Precompensation and Modelling Power Consumption

The precompensation used in the experiments of Section 4.2 did not lead to ideal system performance, as evidenced by the need to use additional receiver-side equalisation for 50 Gb/s signalling. In this section, optimal precompensation with matched filters will therefore be designed. Initially, the effect of precompensation on a simple RC circuit will be evaluated, while also paying attention to the increase in voltage needed for the precompensation, in addition to the resulting

increase in energy dissipated in the circuit.

4.3.1 Voltage and Current Behaviour in an RC circuit

In the following analysis, the values used in the RC circuit will be 350 fF and 50Ω , which represent a moderately doped ($5 \times 10^{17} \text{ cm}^{-3}$) PN phase shifter at -1 V bias being driven by a standard 50Ω source. The junction resistance R_j is ignored for now due to being much smaller than the source resistance.

The derivation of the frequency responses of the current drawn $I(\omega)$ and voltage $V_c(\omega)$ dropped across the capacitor for an RC circuit (derived in Appendix B) are shown, respectively, below:

$$H_i(\omega) = \frac{1}{\sqrt{R^2 + 1/(\omega C)^2}} \cdot e^{i\phi(\omega)}, \quad (4.1)$$

$$H_{vc}(\omega) = \frac{1}{\sqrt{1 + (\omega RC)^2}} \cdot e^{i\left(\phi(\omega) - \frac{\pi}{2}\right)}. \quad (4.2)$$

The time domain representations of these expressions are easily obtained by suitable application of inverse fast Fourier transform (IFFT), i.e., $h_i(t) = \text{IFFT}\{H_i(\omega)\}$ and $h_v(t) = \text{IFFT}\{H_v(\omega)\}$. For a single-toned, sinusoidal waveform, the power consumed by the circuit can be calculated in the frequency domain by

$$P(\omega) = \text{Re}\{V_s(\omega) \times I^*(\omega)\}, \quad (4.3)$$

where I^* represents the complex conjugate of the current. For PRBS waveforms, which contain a spread of frequencies, the power consumed can be calculated by integrating Equation 4.3 over the relevant frequencies, or else obtained using a dot product in the time-domain by

$$P(t) = \frac{1}{2} \text{Re}\{V_s(t) \cdot I^*(t)\}. \quad (4.4)$$

These equations for the frequency response and power dissipation can now be used to simulate the performance of MZM that follows an RC frequency roll-off. The stages of such a model are shown in Figure 4.9.

The frequency response (in terms of magnitude squared) of the voltage falling across the capacitor is shown against frequency in Figure 4.10(a). Also shown for comparison is the corresponding response of the voltage across the resistor (equivalent to the reciprocal of the first curve). The 3 dB point of the capacitor's

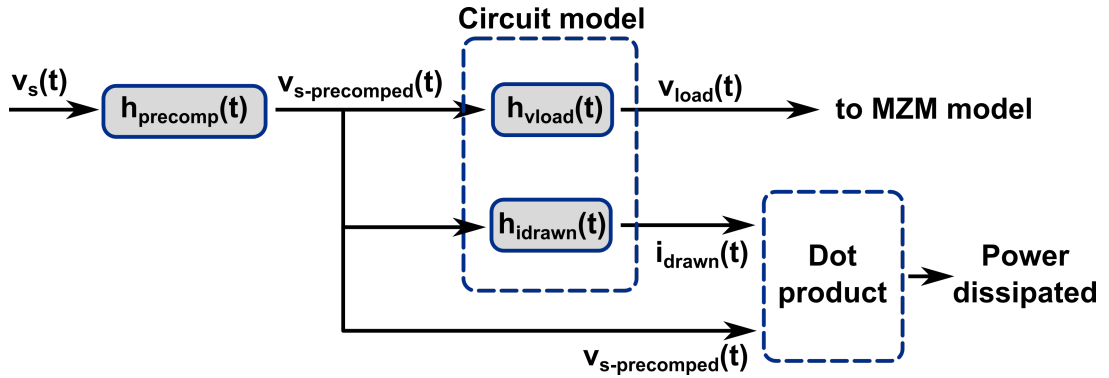


Figure 4.9: Flow chart describing the stages of the circuit model.

frequency response occurs when the the voltages across the resistor and capacitor are equal. Clearly, as the frequency increases, the resistive element of the circuit starts to dominate. This explains the plot of total power dissipated in the circuit for a 1 V peak-to-peak sinusoidal drive voltage in Figure 4.10(b). The power dissipation increases with frequency and begins to look largely resistive at high frequencies, approaching V^2/R .

However, it is not the power that is the main metric of interest when it comes to energy consumption; the energy consumed per bit in the circuit, which represents the efficiency of the power use, is of more concern. As a stepping stone to this, Figure 4.10(c) shows the energy consumed “per cycle”, i.e., the energy consumed in completing one period of a sinusoidal waveform. This serves as a single-toned analogy to the energy per bit that is normally used for PRBS waveforms. Notably, the energy per cycle peaks at the 3 dB point of the RC circuit before dropping off. Beyond this point, as the power consumption changes from capacitive in nature (which depends on switching speed, and, hence, frequency) to resistive (which does not depend on frequency), the energy use per cycle starts to become more efficient at higher frequencies. Finally, building on this, Figure 4.10(d) shows the energy dissipated per bit in the RC circuit for NRZ waveforms (PRBS-15) of various symbol rates. As the NRZ spectrum is spread out over frequencies as a sinc function, the energy consumed per bit is essentially an integrated version of Figure 4.10(c) and so largely flattens out at higher symbol rates. The typical expression that is normally used to estimate capacitive, lumped modulator energy consumption was given in Equation 2.28, and the output of this is also shown in Figure 4.10(d) for comparison. It gives a reasonable estimation of the energy consumption (largely within a factor of 2), but it is clear that the frequency responses of the various elements in the circuit do play a role in the overall energy dissipation. Most notably, at higher frequencies, the shift of voltage drop from

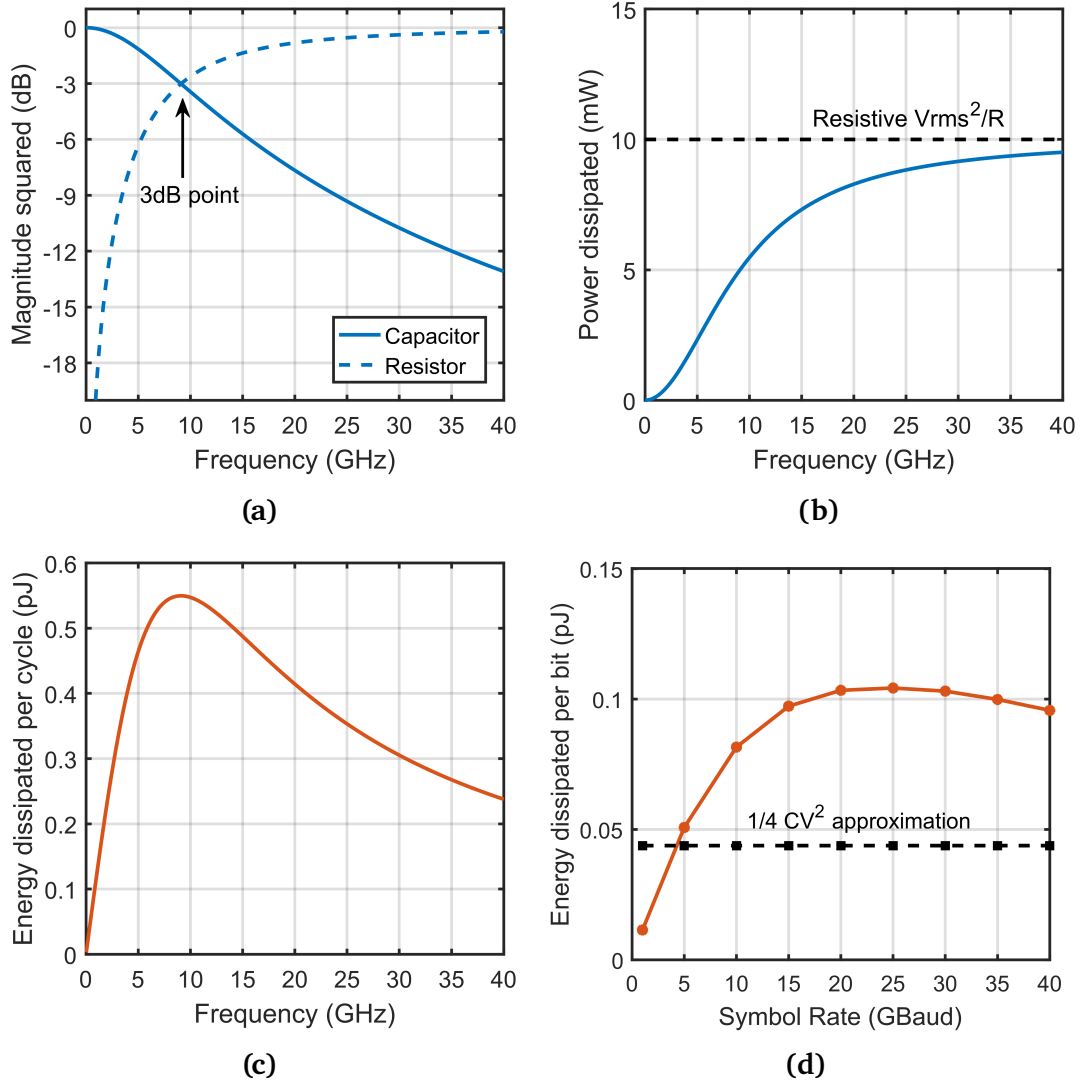


Figure 4.10: (a) Frequency response of the resistor and capacitor in the RC circuit, where $R = 50 \Omega$ and $C = 350 \text{ fF}$. (b) Total power dissipated for an applied sinusoidal voltage. (c) Energy dissipated per cycle of the sinusoid. (d) Energy dissipated per bit for an applied PRBS waveform.

the capacitor to the resistor causes additional, significant power consumption in the circuit.

4.3.2 Precompensation for an RC circuit

As a simple design of precompensation, we can investigate boosting the 3 dB bandwidth of the RC circuit by a given factor, which we will call N_{3dB} . From Equation 2.22, we recognise that it is the RC time constant that determines the 3 dB bandwidth of the RC circuit. Hence, the time constant must be scaled by a factor of $1/N_{3dB}$ to give the required boost in bandwidth. By inspection of

Equation 4.2, we see that scaling the amplitudes of the drive waveforms with the following precompensation filter will give the desired frequency response:

$$H_{precomp}(\omega) = \frac{\sqrt{1 + (\omega RC)^2}}{\sqrt{1 + (\omega RC/N_{3dB})^2}}. \quad (4.5)$$

After an inverse Fourier transform, the resulting precompensation taps can be applied to waveforms in the time domain. Figure 4.11(a) shows the energy dissipated per bit in the circuit by NRZ waveforms with various values of the boost factor ranging from 1 to 2.5 applied. The RC circuit in question has a 3 dB bandwidth of 9.1 GHz, so these boost factors result in corresponding boosted bandwidths ranging from 9.1 GHz to 22.7 GHz. As no normalisation of the taps has been made, they give gain to the higher frequencies in the drive waveform rather than attenuating the low frequencies, thus raising the RMS value of the voltage. Because of this, the energy consumption naturally rises both with symbol rate and choice of boost factor. Figure 4.11(b) emphasises the increased energy demand of the precompensation by plotting the energy per bit for only the 25 Gbaud NRZ waveform against N_{3dB} . This shows that the required energy per bit doubles as the boost factor is varied from 1 to 2.5.

When precompensating for low-bandwidth, however, it is important to evaluate the overall effect on system performance. After all, a boost factor of 2.5 has doubled the energy per bit, and increased the RMS voltage of the driver by a factor of 2.5. We must therefore ask if the improvement in signal quality is worth the effort. Inset in Figure 4.11(b) are example 25 Gbaud eye diagrams resulting from boost factors of 1, 2 and 2.5. Clearly, the eye opens significantly as the bandwidth is increased. Examining the resulting improvement in receiver sensitivity for the various boost factors in Figure 4.11(c), the benefit is seen to be rather modest, with just over 1 dB improvement for a boost factor of 2.5. Also shown for comparison is the sensitivity improvement needed to match the performance of an “ideal” MZM, with a frequency response that follows a fourth-order Bessel filter with 3 dB bandwidth of 25 GHz. No matter the choice of N_{3dB} , the sensitivity of the bandwidth-boosted RC circuit never reaches that of this ideal case, with a slight difference of approximately 0.5 dB. The reason for this residual penalty is that the chosen precompensation only deals with the magnitude of the RC circuit’s roll-off, and not the phase. The phase of the RC circuit still has a frequency dependence (derived and plotted in Appendix B) that

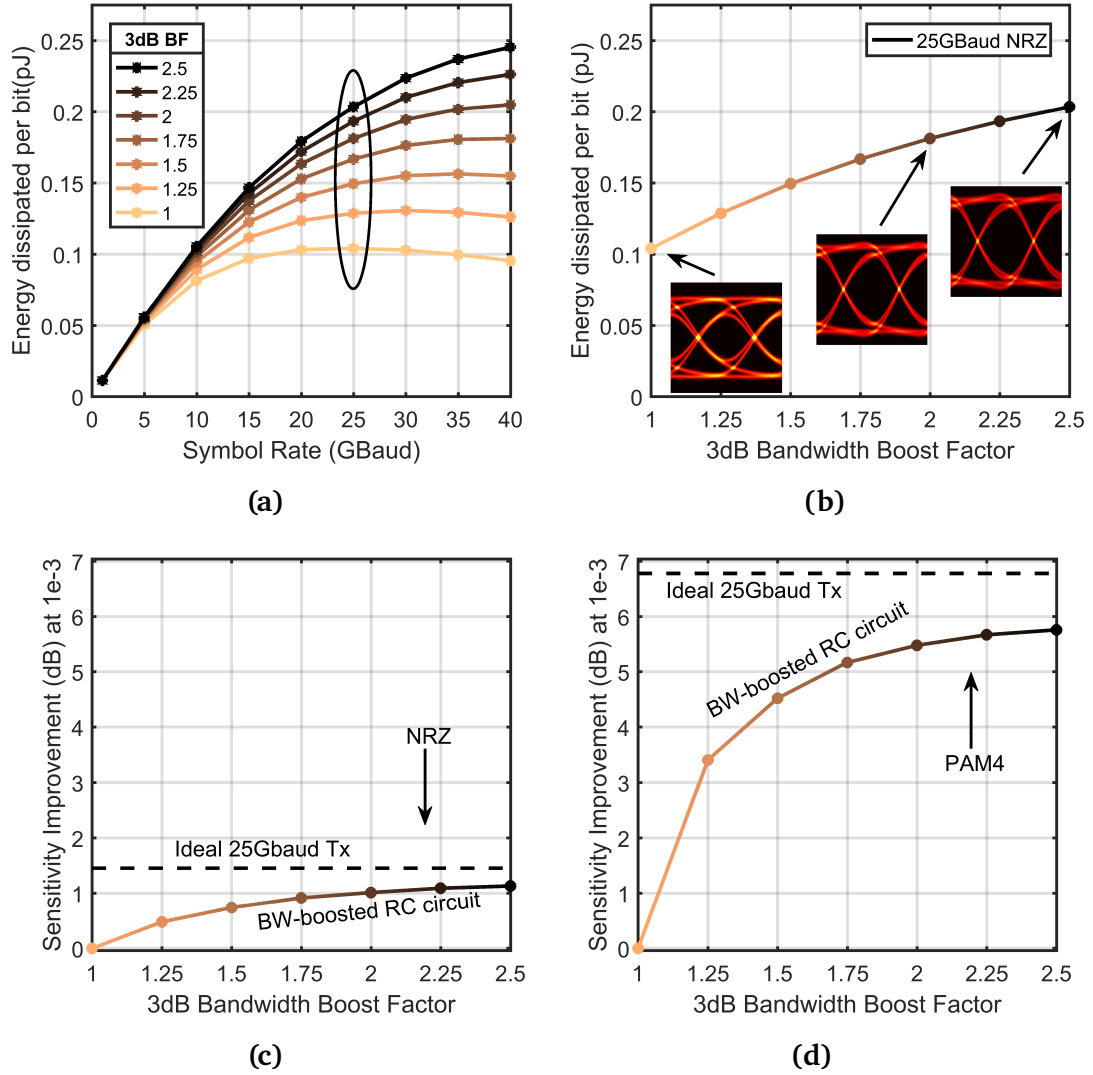


Figure 4.11: (a) Energy dissipated per bit in the RC circuit for various bandwidth boost factors shown against NRZ symbol rate. Here, $R = 50 \Omega$ and $C = 350 \text{ fF}$. (b) A focus on energy dissipated for 25 Gbaud NRZ against boost factor, with inset corresponding eye diagrams. (c) Sensitivity improvement for 25 Gbaud NRZ against boost factor. (d) Corresponding sensitivity improvement for 25 Gbaud PAM4.

is given by

$$\phi(\omega) = \tan^{-1} \left(\frac{1}{\omega RC} \right). \quad (4.6)$$

As this phase does not have a linear dependence on frequency, the resulting non-flat group delay negatively affects the signal quality. This is visible as slight overshoot and undershoot in the NRZ eye diagram.

Although the improvements in receiver sensitivity seem modest for 25 Gbaud NRZ, the value of precompensation is made more clear in Figure 4.11(d) where the corresponding sensitivity improvement against bandwidth boost factor is shown for 25 Gbaud PAM4. With PAM4 being less resilient to eye closure than NRZ at the same symbol rate, the boosting of the bandwidth by a factor of 2.5 gives nearly 6 dB improvement in receiver sensitivity. There is still residual penalty from the non-flat group delay, which is slightly worse (1 dB penalty) than for the NRZ case. The large improvement in signal quality for PAM4 points to the importance of precompensation when using multilevel signalling in bandwidth-limited systems. The remaining penalty due to phase or group delay also emphasises the need to design an ideal precompensation filter, particularly when the phase and magnitude of an MZM's frequency response can be worse than that of a simple RC circuit. The following section will describe how to design an ideal, matched filter to precompensate for arbitrary magnitude and phase variations.

4.3.3 Determining a Matched Precompensation Filter for Arbitrary Waveforms

We begin by letting $H_{channel}(\omega)$ be the frequency response of the channel, $H_{desired}(\omega)$ the desired frequency content of the signal at the receiver, and $H_{precomp}(\omega)$ the required precompensation filter to obtain this desired response. In the frequency domain,

$$H_{precomp}(\omega) \times H_{channel}(\omega) = H_{desired}(\omega). \quad (4.7)$$

Then the desired precompensation filter is determined by simple element-by-element division through $H_{precomp}(\omega) = \frac{H_{desired}(\omega)}{H_{channel}(\omega)}$. If the desired response is to be flat across all frequencies in the range of interest, then $H_{precomp}(\omega)$ is essentially the reciprocal $H_{channel}(\omega)$, and is known as a zero-forcing filter. This type of precompensation was used for the transmission experiments with

PAM4 waveforms MZM in Section 4.2. However, zero-forcing filters are not necessarily an optimal choice of precompensation, particularly when there is heavy bandwidth roll-off in the system; boosting of heavily-attenuated higher frequencies can introduce noise. Hence, it is better to design a precompensation filter that boosts only the necessary frequencies for the signal, i.e., so $H_{desired}(\omega)$ exactly follows the frequency spectrum of an ideal NRZ or PAM4 signal. The resulting precompensation filter is known as a matched filter.¹

To obtain the matched filter, we start with the time-domain representation of Equation 4.7, which is

$$h_{precomp}(t) * h_{channel}(t) = h_{desired}(t), \quad (4.8)$$

where the element-by-element multiplication has now changed to a convolution due to the inverse Fourier transform involved in moving from frequency domain to time domain. Convoluting both sides of Equation 4.8 with an ideal NRZ² waveform $v_{NRZ}(t)$ and using associativity, we arrive at

$$\underbrace{h_{precomp}(t)}_{w(t) \text{ To be determined}} * \underbrace{[v_{NRZ}(t) * h_{channel}(t)]}_{x(t) \text{ Distorted waveform}} = \underbrace{v_{NRZ}(t) * h_{desired}(t)}_{y_{target}(t) \text{ Desired waveform}}. \quad (4.9)$$

This distorted waveform $x(t)$ is what is typically captured at the receiver end of the system (through an oscilloscope), and the desired waveform $y_{target}(t)$ is already known. This resulting convolution operation is implemented digitally through a matrix operation, an example of which is shown below in Equation 4.10 where 10 taps weights are used to precompensate a waveform that is 10 000 samples long:

$$\begin{bmatrix} w_1 & w_2 & \dots & w_{10} \end{bmatrix} \begin{bmatrix} x_1 & x_2 & \dots & x_{10001} \\ x_2 & x_3 & \dots & x_{10002} \\ \vdots & \vdots & \ddots & \vdots \\ x_{10} & x_{11} & \dots & x_{10010} \end{bmatrix} = \begin{bmatrix} y_1 & y_2 & \dots & y_{10000} \end{bmatrix}. \quad (4.10)$$

This is a system of linear equations, the solution of which gives the desired set

¹The desired response may also be convolved with an ideal bandwidth roll-off to reduce high-frequency noise at the transmitter. In this work, a fourth-order Bessel with 3 dB bandwidth equal to 1 or 0.75 times the symbol rate is typically used.

²An NRZ waveform is used here, but the treatment for a PAM4 or EDB waveform is no different.

of precompensation taps $h_{precomp}(t)$. There exist many methods of finding this solution. Here, we use Matlab's *mldivide* function, which is a convenient general solver for linear equations.

4.4 Designing Optimal Precompensation for the Lumped MZM

In Section 4.2, when carrying out the experimental measurements of 50 Gb/s PAM4 and EDB, additional receiver-side (Rx) equalisation was necessary to achieve bit error rates below the required FEC thresholds. As the precompensation used at the time was not optimal, the Rx equalisation compensated for residual ISI that was present in the system. Now that a suitable method of obtaining matched precompensation filters has been determined, 50 Gb/s transmission with PAM4 and EDB using the lumped MZM can be re-examined to verify that Rx equalisation is not needed.

At this point, a discussion of reflections in the RF cables connected between the DAC and lumped MZM in the experimental setup becomes important. While attempting to train the matched filters to compensate the measured frequency response of the lumped MZM, it was noted that the filter lengths (in the time domain) that were needed for an optimal signal were very long: with hundreds of taps required. For a 100 GS/s DAC, this corresponds to times in the order of nanoseconds having an influence in terms of ISI. This is unusual, as it implies that extremely low-frequency characteristics of the response were having a major effect on the signal quality.

Afterwards, on further investigation, it was noticed that the moving-average smoothing that was applied digitally to the frequency response to deal with measurement noise was hiding a key feature—there were large ripples present in the frequency response. Such ripples are a typical indicator of reflections in the RF cables of the drive setup. The period of the ripples in the frequency spectrum corresponded to a Fabry-Perot resonant structure with a round-trip time of 5.2 ns. This suspicion was confirmed by looking at the impulse response of the measurement, where a small “bump” or extraneous peak was noticed at a delay of 5.2 ns from the main peak. These features are shown in the frequency domain and time domain in Figure 4.12(a) and Figure 4.12(b), respectively. Noticeably, even though this reflection appears relatively small in the impulse response (with a relative linear amplitude of approximately 0.03), its effect in

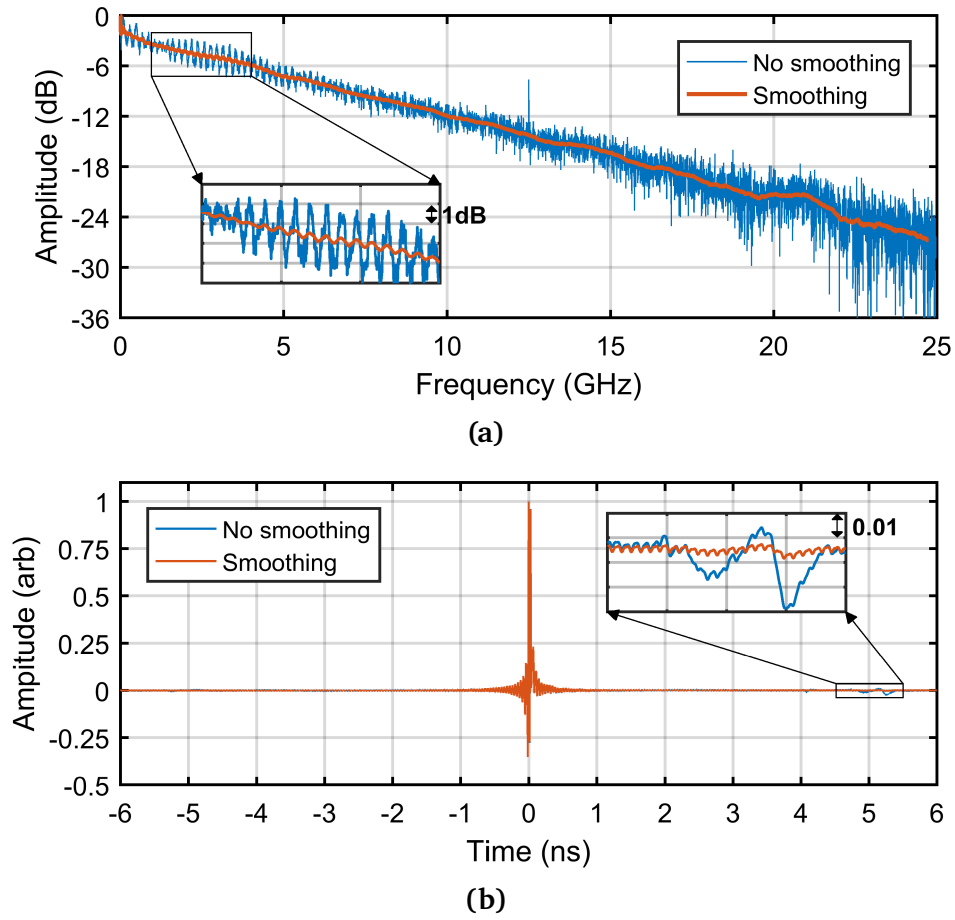


Figure 4.12: Measured (a) frequency response and (b) impulse response of the lumped MZM, with and without smoothing. The zoomed, inset sections highlight how smoothing can mistakenly conceal RF reflections in the system.

the frequency domain is significant (causing a peak-to-peak variation of up to 3 dB.)

Using a typical velocity factor (74 %) for the 2.92 mm K-type RF cables in the setup [136], this delay corresponds to travelling a distance of 1.15 m. On reviewing the experimental setup, this was pinpointed to the round-trip time in a cable of length 0.6 m that connected the RF amplifier to the lumped MZM. This implies that the reflection was happening between the output of the amplifier and the lumped MZM. This is a reasonable conclusion, given the significant impedance mismatch between the $50\ \Omega$ amplifier output and the quasi-open-circuit presented by the lumped MZM.

This reflection effect can be removed in simulations of the channel by flattening out the bump, i.e., simply setting the filter coefficients of the measured impulse response in the region of the reflection to zero, or truncating the impulse response

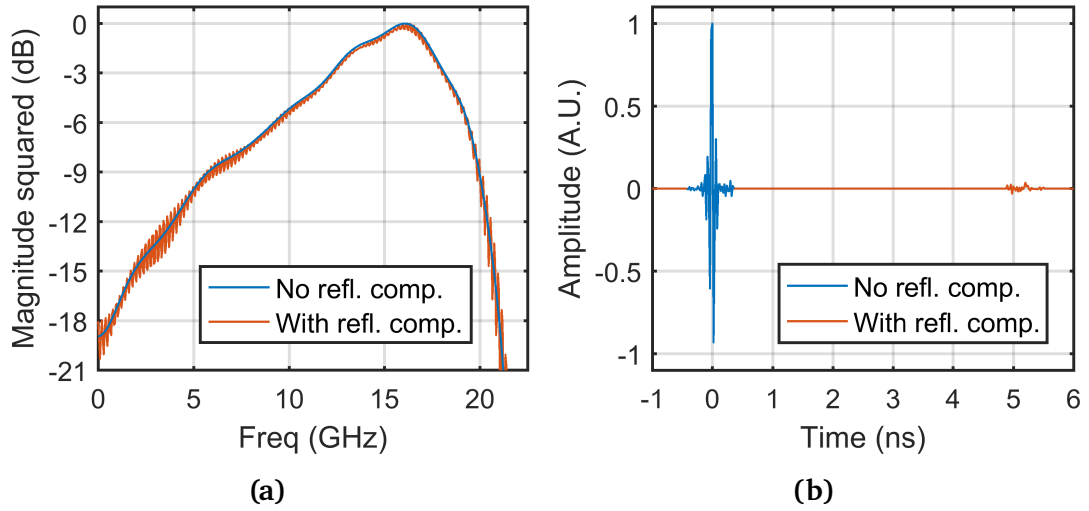


Figure 4.13: Comparison of the matched precompensation filter for 25 Gbaud PAM4 in terms of (a) frequency response and (b) impulse response, with and without the additional taps to compensate for the RF reflections in the drive cables. Note the “bump” present in the time domain near 5 ns.

before this point. However, to compensate for it in experimental measurements (in which precompensation is being used), no smoothing must be applied to the measured frequency response of the channel, and it must be ensured that a sufficient quantity of filter taps are used so that they span the reflection in the time domain. For example, consider the situation where a fractionally spaced, matched FIR filter is to be used to equalise a 25 Gbaud PAM4 signal to be sent by the lumped MZM. Figures 4.13(a) and 4.13(b) compares the frequency response and impulse response of this matched filter when an insufficient number (80) and a sufficient number (800) of taps are trained. In contrast with the shorter filter, the longer filter clearly exhibits the necessary ripples in its frequency response to compensate the RF reflection.

Compensating the reflections during the experimental measurements in this manner does not give a loss of generality to the analysis for two reasons. Firstly, the reflections arise as a result of the output impedance of the discrete RF amplifier used in this setup. It can be assumed that impedance matching can be improved in optimal design of the amplifier used in a driver chip for the MZM. Secondly, the reflections are more noticeable due to the length of the RF cable present after the amplifier; the electrical lengths involved in a flip-chipped or wire-bonded driver would be orders of magnitude shorter, and would reduce the severity of the reflections. As an example, assuming a wire bond length of 0.5 mm, the time delay for a reflected pulse would be in the order of ps, and

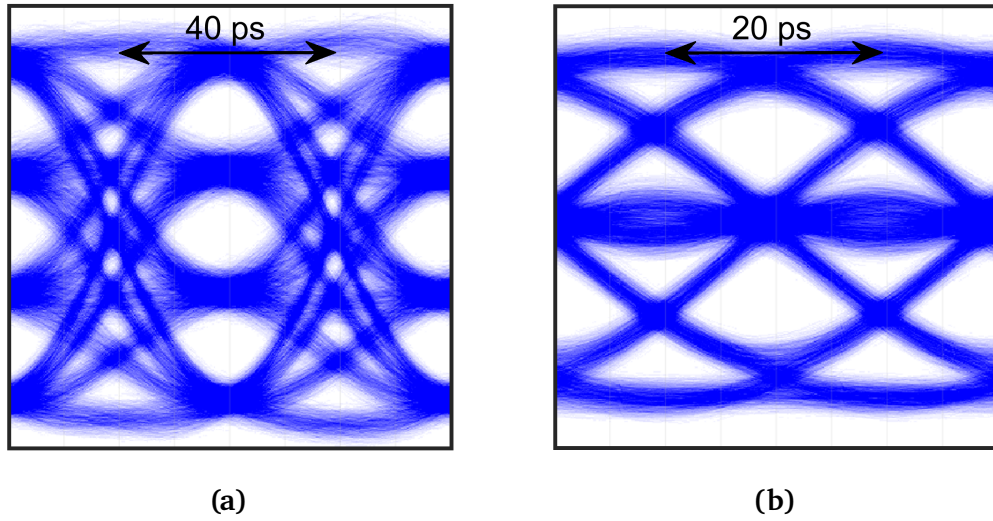


Figure 4.14: Measured 50 Gb/s eye diagrams for (a) PAM4 and (b) EDB waveforms obtained from the lumped MZM with matched precompensation filter. PAM4 ER is 3 dB. EDB ER is 3.3 dB. The symbol period is annotated on each.

would therefore be subsumed into the impulse response of the MZM.

With a suitable method determined for better precompensation of the lumped MZM, we can return to examining its performance for 50 Gb/s systems. For this, the same setup as Figure 4.2 was used. Only the lumped MZM was investigated here, as it is a more typical example of a lumped design than the unterminated TW-MZM. To measure the ER of the MZM's output when no precompensation is applied, a lower-speed signal must be sent to avoid the MZM's severe bandwidth limitation causing full eye closure and affecting the measurement. As the implementation of the DAC in the drive setup had a lower bound on its possible baud rates, a separate, lower-speed pulse-pattern generator (PPG) was connected instead to the RF amplifiers to drive the MZM with a low-speed 5 Gbaud NRZ signal. The ER for this non-precompensated case was calculated from the resulting optical signal's rails and found to be 9 dB for a peak-to-peak drive voltage of 3 V in each arm. The drive setup was then switched back to use the high-speed DAC rather than the PPG, and 50 Gb/s PAM4 and EDB signals with matched-filter precompensation were then transmitted. The peak-to-peak drive voltage into each arm of the MZM was again 3 V. Figure 4.15(a) and Figure 4.15(b) show, respectively, the resulting eye diagrams when measured in a back-to-back link. The optical input power to the receiver is -5 dBm in both eyes. The impact of precompensation on ER is similar for both modulation formats, with the ER dropping to 3 dB for the PAM4 signal and 3.3 dB for the

EDB signal. Both eyes exhibit clear openings, and appear free of any significant bandwidth limitations.

To subsequently investigate the performance of the lumped MZM when transmitting over lengths of fibre, it was decided to switch the analysis method to a simulation testbed using the measured frequency response of the device. This allowed for a high-resolution sweep of the fibre length while removing unnecessary experimental variables. The code for the testbed was internally developed and written entirely in Matlab. The pattern generation and error analysis functions used are identical to those in the experimental setup of Figure 3.29; the simulation testbed simply replaces the DAC, laser source, MZM, optical fibre, and photoreceiver with suitable models. The MZM model follows Equation 3.9, where the phase shift is provided by 1.5 mm long silicon phase shifters. The junctions of these phase shifters have equal doping P and N-type doping of $5 \times 10^{17} \text{ cm}^{-3}$, and are biased at -1 V and modulated with peak-to-peak voltages of 1 V . The measured frequency response of the lumped MZM is applied electrically at the transmitter to give the same bandwidth limitation as experimental testbed. The modelled fibre is conventional SMF, with a dispersion parameter of 17 ps/nmkm and attenuation of 0.2 dB/km in the C-band. The modelled photodiode has an NEP of $24 \text{ pW}/\sqrt{\text{Hz}}$ and an ideal responsivity of 1.

The resulting BER curves for 50 Gb/s PAM4 and EDB over increasing lengths of fibre are shown, respectively, in Figure 4.15(a) and Figure 4.15(b). The PAM4 system has worse sensitivity than the EDB in a back-to-back link, crossing the DH-FEC and KP4-FEC thresholds at -10.1 dBm and -8.5 dBm , respectively, as opposed to -11.5 dBm and -9.75 dBm for EDB. However, PAM4 is seen to be more resilient than EDB as the fibre length increases due to its more compact spectrum; the 10 km distance required for the LR8 400G standard can be comfortably reached using KP4-FEC for PAM4, while EDB requires the more heavy-duty HD-FEC to just about reach the same distance. The effect of transmission distance on performance is shown more clearly in Figure 4.15, where the receiver sensitivity at the KP4-FEC threshold is plotted against fibre length. EDB outperforms PAM4 by approximately 1.5 dB over short distances, but is less suitable as the distance increases past 6 km. These results (for transmitter-side equalisation) show a similar trend to a comparison of 50 Gb/s PAM4 and EDB in the literature that used solely receiver-side equalisation [137] to compensate heavily bandwidth-restricted transceivers. Due to the need for data centre transceivers to reach distances of 10 km (and possibly longer), the

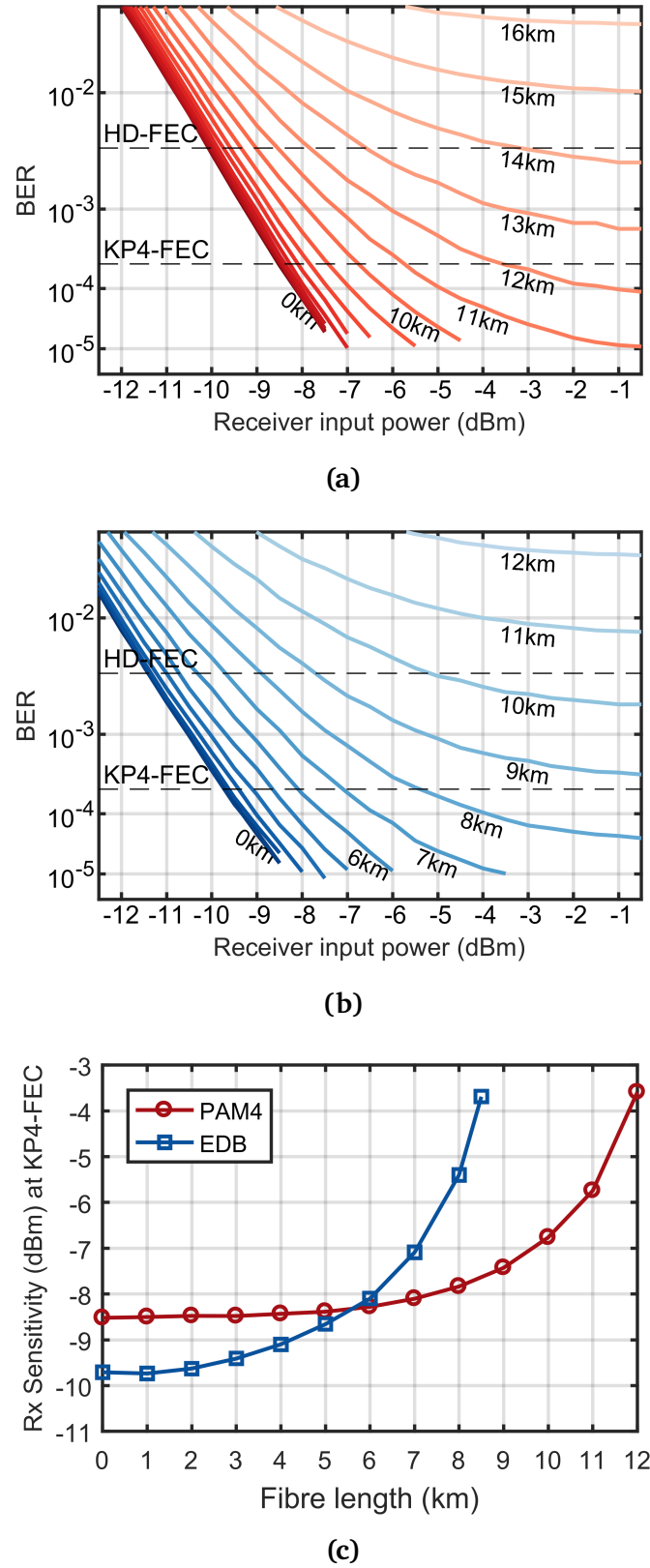


Figure 4.15: Simulated BER curves over various distances for (a) PAM4 and (b) EDB from the precompensated lumped MZM. (c) Receiver sensitivity at the KP4-FEC threshold.

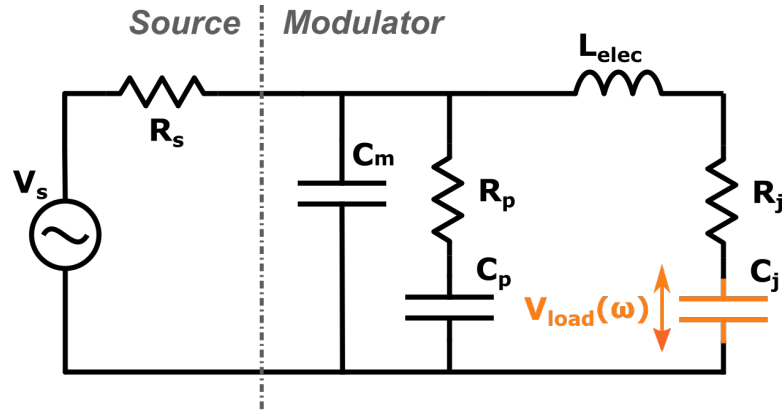


Figure 4.16: Equivalent circuit of the driver and phase shifter for a lumped MZM, including parasitics.

remainder of the analysis in this chapter will focus on using PAM4 signalling with lumped MZMs rather than EDB.

4.5 Modelling the Frequency Response of the Lumped MZM

In Section 2.5, the energy per bit for a lumped MZM was calculated under the assumption that the typical $\frac{1}{2}CV^2$ expression held. This expression is suitable for the low-frequency regime, where the capacitor in question (the PN junction phase shifter) charges and discharges fully. However, at high frequencies, a more thorough treatment of the circuit is required. The following section derives a suitable model for the frequency response of the lumped MZM, including parasitic elements that expand the treatment of its bandwidth beyond that of a simple RC circuit. Attention is also paid to the effect of the optical transit time of the light through the MZM on its bandwidth.

4.5.1 Deriving an Accurate Equivalent Circuit Model of a Lumped MZM from S11 Measurements

In order to accurately model the bandwidth and energy consumption of a lumped MZM, a suitable equivalent circuit model must be obtained. Typically, some variety of LRC circuit is used to model lumped devices of this type [101, 138]. Figure 4.16 shows such a circuit, based on a model provided by the foundry for this lumped MZM (IMEC) [139]. Here, R_s is the source resistance of the driver (typically $50\ \Omega$), C_m is the parasitic capacitance of the RF bond pad, R_p is

the resistance of the bulk silicon, C_p is the capacitance through the oxide layer, and L_{elec} is the inductance of the electrodes. The parameters R_j and C_j are the junction resistance and capacitance, respectively, which are dependent on the doping levels of the silicon phase shifter as discussed in Section 3.3.3. For a given angular frequency ω , the complex impedances of these parameters can be obtained in the usual fashion: the complex impedance for a resistance R is given simply by R , while the complex impedances of a capacitance C and inductance L are given by $\frac{1}{i\omega C}$ and $i\omega L$, respectively. From here on, the complex impedances resulting from the circuit parameters in Figure 4.16 will be respectively labelled as Z_{Rs} , Z_{Cm} , Z_{Rp} , Z_{Cp} , Z_{Lelec} , Z_{Rj} , and Z_{Cj} . As the voltage dropped across the junction capacitance is the crucial load voltage for obtaining modulation with the MZM, this has been labelled $V_{load}(\omega)$.

To model the values of the resistances, capacitances and inductance that fit the equivalent circuit for a range of frequencies, a suitable parameter must be found that can be both measured experimentally and calculated analytically using all the components of the circuit model. Numerical fitting methods can then be used to derive the values of the various circuit elements. The optimal parameter for such a task is the reflection parameter, also known as the S11 parameter, of the device. This is a measurement of the complex ratio of the voltage of the reflected wave (V^-) from the lumped MZM to the voltage of the incident wave (V^+) from the source, typically written as $\Gamma = \frac{V^-}{V^+}$. In terms of impedances, it can also be expressed as

$$\Gamma = \frac{Z_{mod} - Z_{source}}{Z_{mod} + Z_{source}}, \quad (4.11)$$

where Z_{mod} and Z_{source} are the complex impedances of the MZM chip and voltage source, respectively. As the silicon phase shifters are PN junctions, an unterminated, lumped MZM behaves like a diode; we can expect to see a very high reflection parameter when the MZM is under reverse bias.

A VNA can be used to measure the S11 parameter over a range of frequencies. In this setup, a K-cable connects the VNA to a high-speed GS probe, which then contacts a ground pad and signal pad of one arm of the MZM. As some circuit parameters are expected to be very small, it is crucial that the effect of all measurement equipment between the VNA and MZM (i.e., the cable and probe) is first calibrated out to avoid disturbing the accuracy of the fit. For this, a Picoprobe CS-8 calibration substrate is used. After measurements are made of the S11 parameter of a standard short (S), open (O) and 50Ω load (L) on the substrate, with the correct calibration constants for the probe and substrate

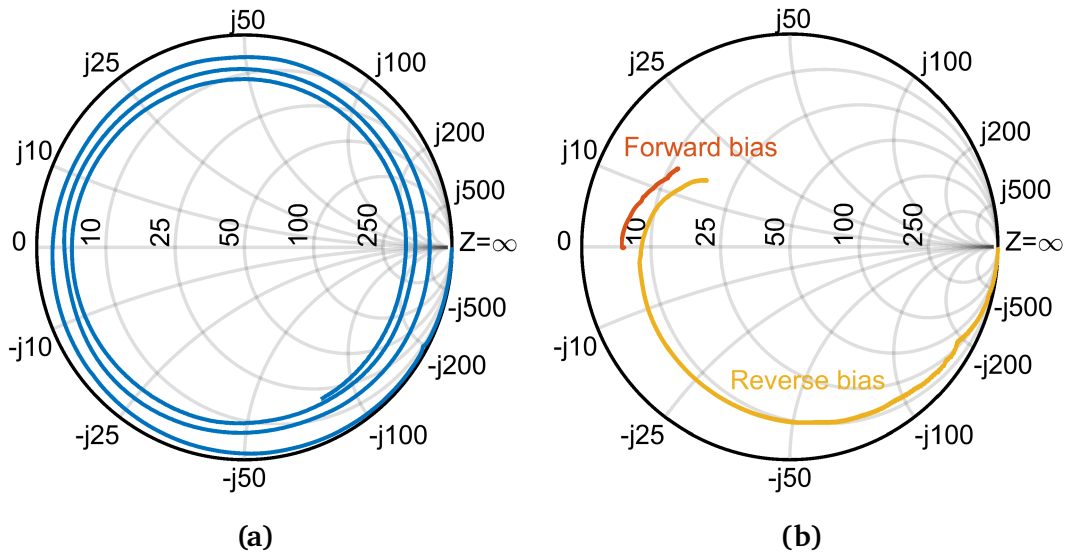


Figure 4.17: (a) Smith chart plot of the measured S_{11} of a reference unloaded transmission line after calibration. (b) Comparison of the measured S_{11} of the lumped MZM under forward bias (1 V) and reverse bias (-1 V). All measurements run from 10 MHz to 40 GHz.

pair inputted, the VNA carries out an SOL calibration to de-embed the cable and probe from subsequent S_{11} measurements of the MZM. To visually evaluate the calibration, Figure 4.17(a) shows a Smith chart for an S_{11} measurement from 10 MHz to 40 GHz of an unloaded, long reference transmission line on the calibration substrate, after calibration has been carried out. The S_{11} parameter winds clockwise from full reflectance (the $Z = \infty$ point) at low frequency, remaining smooth and close to the outside of the chart, which points towards a good calibration. As the frequency increases, the S_{11} parameter starts to move slightly inwards, as expected, due to increased RF losses in the transmission line.

Figure 4.17(b) plots, on a Smith chart, the S_{11} parameter of the lumped MZM under two example conditions: reverse bias (-1 V) and forward bias (1 V). Under reverse bias, the MZM behaves quite similarly to the open transmission line, with the S_{11} parameter showing full reflectance at low frequencies before winding gradually inwards as the frequency increases. In contrast, for forward bias, the MZM's S_{11} parameter starts on the other side of the Smith chart, closer to a short circuit. As mentioned above, this diode-like behaviour is expected given the PN junction phase shifters in the arms.

To examine the S_{11} parameter further, Figure 4.18(a) and Figure 4.18(b) show the measured magnitude and phase of the S_{11} parameter, respectively, for various reverse biases. Again, the lumped MZM shows full reflection of the

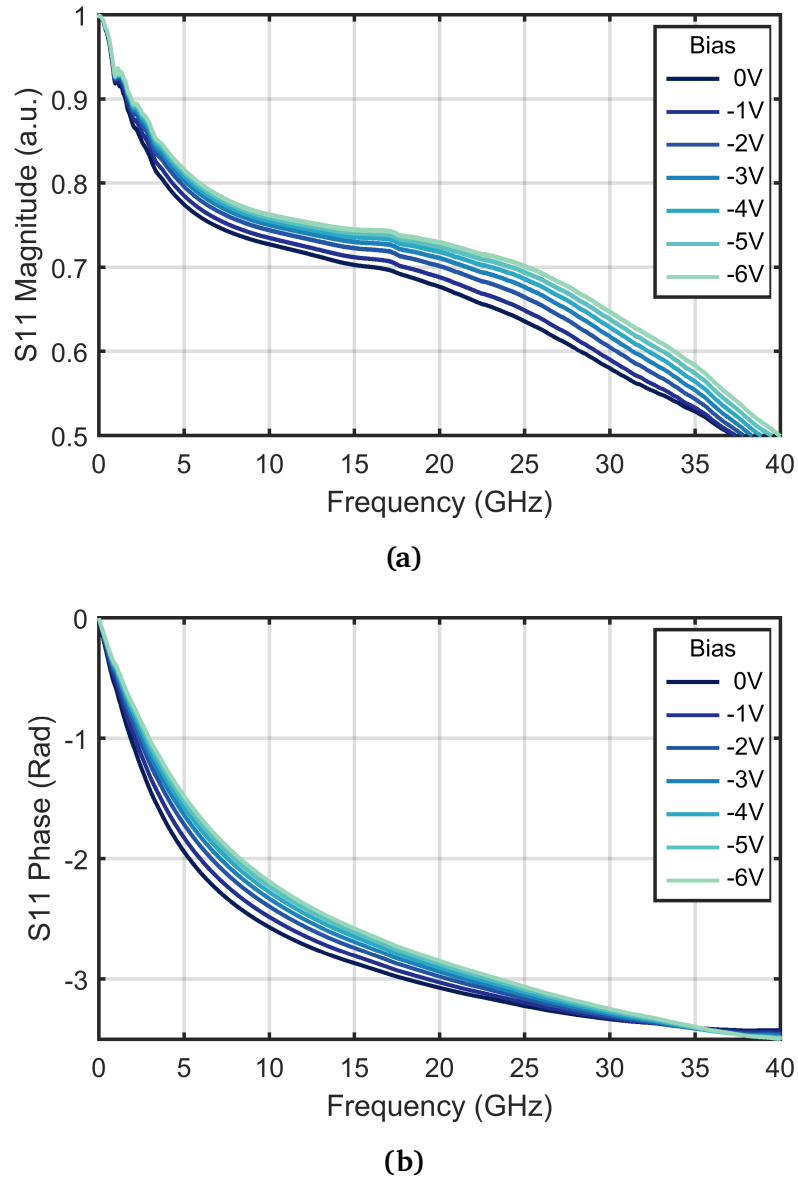


Figure 4.18: Measured S11 parameter of the lumped MZM in terms of (a) magnitude and (b) phase.

VNA input signal at low frequencies. Essentially, the MZM presents an infinite impedance initially as its behaviour is dominated by the capacitance C_j of the PN junction. Due to the presence of parasitic resistance and inductance in the circuit, the magnitude of the S11 parameter then decreases as frequency increases. Increasing the reverse bias results in a more gentle roll off in magnitude. This is due to C_j decreasing with bias, causing the magnitude of the corresponding complex impedance Z_{Cj} to increase for all frequencies, presenting a greater impedance mismatch to the $50\ \Omega$ VNA source.

In order to fit the equivalent circuit of the lumped MZM to the measured

S11 parameter, the reflection coefficient between the source and modulator in Figure 4.16 must be first derived. By summing appropriately the complex impedances in series and parallel, the impedance of the modulator Z_{mod} can be expressed as

$$Z_{mod} = Z_{Cm} \parallel (Z_{Cp} + Z_{Rp}) \parallel (Z_{Rj} + Z_{Cj} + Z_{Lelec}). \quad (4.12)$$

The reflection coefficient Γ between source (VNA) and modulator is then calculated for the necessary frequencies by inserting Z_{mod} into Equation 4.11 and setting Z_{source} to 50Ω .

With a suitable expression for the reflection coefficient, the measured S11 parameter can now be fitted. To perform the fitting, we use Matlab's *fmincon* function, which is designed to find the minimum of a constrained non-linear multivariable function. In this case, the function to be minimised is the mean squared error (MSE) between the analytical expression for Γ and the corresponding measured data for frequencies from 10 MHz to 40 GHz (the maximum available frequency provided by the VNA). As Γ is complex valued, its real and imaginary parts must be separated during the fitting; the respective differences between the real and imaginary parts of the measured data and analytical expression are calculated individually before being combined into one MSE. As there are a number of circuit elements to be fitted, it is important to constrain their values with upper and lower bounds. This prevents any parameter surging to an unreasonable value in the minimisation process. Appropriate starting values and bounds for the fit are decided upon using sample circuit values provided by IMEC for a shorter 0.5 mm MZM in their iSIPP25G process [139], combined with coarse brute searches over the potential parameter spaces to identify trends in the fit. For the final fittings across the range of bias voltages, the values of L_{elec} , C_m , R_p and C_p are held fixed at sensible mean values obtained from trial fits to reflect the fact that the electrode, bond pads and bulk silicon structure should be unaffected by changes in voltage.

The training curves produced using the final *fmincon* fitting routine are shown for each of the bias voltages in Figure 4.19. The flat line at initial iterations indicates that the starting parameters are, as could be expected, not optimal; the later sharp drop in MSE occurs when the fitting routine has moved the parameters towards a better fit. Each of the curves converges to very low MSE, indicating that a local minimum for the fitting has been found. The training curves for the higher bias voltages start from relatively low MSE values. This is simply due

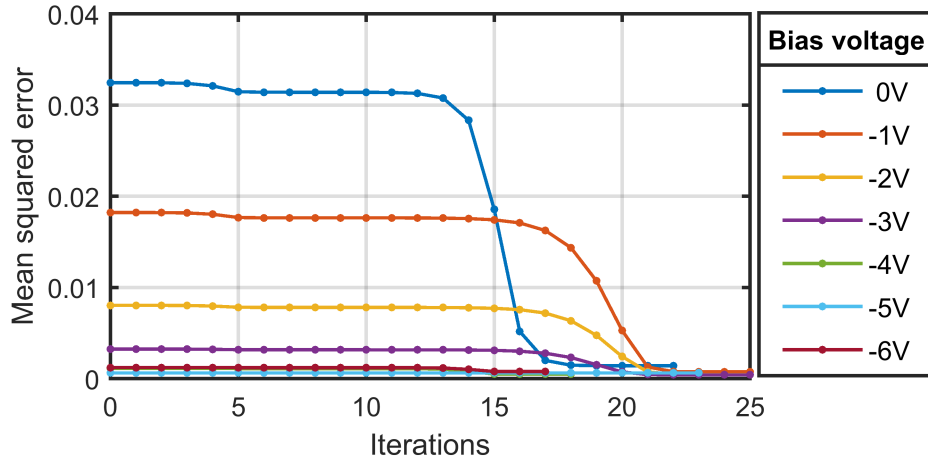


Figure 4.19: Training curves for the final fit of the equivalent circuit model to the measured S11 parameters for various bias voltages.

to the fact that the chosen starting parameters (which are the same for all bias voltages) are quite close to the optimal fit parameters.

The resulting fitted circuit parameters are shown in Table 4.1 for the various reverse bias voltages. The MSE is also shown and is seen to be slightly worse at 0V bias than the other voltages. A potential explanation for this is the square-root dependence of the PN junction capacitance on voltage, which is a more linear relationship at larger bias. We will next discuss the fitted values and compare them with what could be expected from prior knowledge. The fitted value of approximately $10\ \Omega$ for R_j is a little more than twice what would be expected for the phase shifter resistance modelled in Figure 3.17(a). This can be attributed to the skin effect for the reasonably high frequencies in question [101]. The fitted value for C_j is relatively high. As the doping concentrations used in the multi-project wafer fabrication process for this device are not provided, the reason for this cannot be determined exactly. The fitted junction capacitance value of 685 fF at -1 V would place it beyond the range of modelled doping levels shown in Figure 3.16(a), for example, and would be closer to a doping level of $2 \times 10^{18}\text{ cm}^{-3}$. Such a high doping concentration would certainly limit the bandwidth, and so a lower doping concentration and junction capacitance value will be used for the modelling and analysis later in Sections 4.5.4 and 4.6 of this chapter. The fitted value of electrode inductance L_{elec} is reasonably in line with that expected from [139]. So too are the values for bulk silicon resistance R_p and oxide layer capacitance C_p . The capacitance of the bond pad C_m is larger than the typical 10 fF to 30 fF that would be expected [139, 43], which can be attributed to the presence of additional on-chip parasitics.

Table 4.1: Circuit values obtained from fitting of the measured S11 parameter from the lumped MZM for various bias voltages.

Bias [V]	R_j [Ω]	C_j [fF]	L_{elec} [pH]	C_m [fF]	R_p [Ω]	C_p [fF]	MSE []
0	10.7	770	75	100	1000	200	0.0014
-1	10.6	685	75	100	1000	200	0.0007
-2	10.4	612	75	100	1000	200	0.0004
-3	10.3	564	75	100	1000	200	0.0004
-4	10.2	529	75	100	1000	200	0.0005
-5	10.0	501	75	100	1000	200	0.0006
-6	9.9	477	75	100	1000	200	0.0008

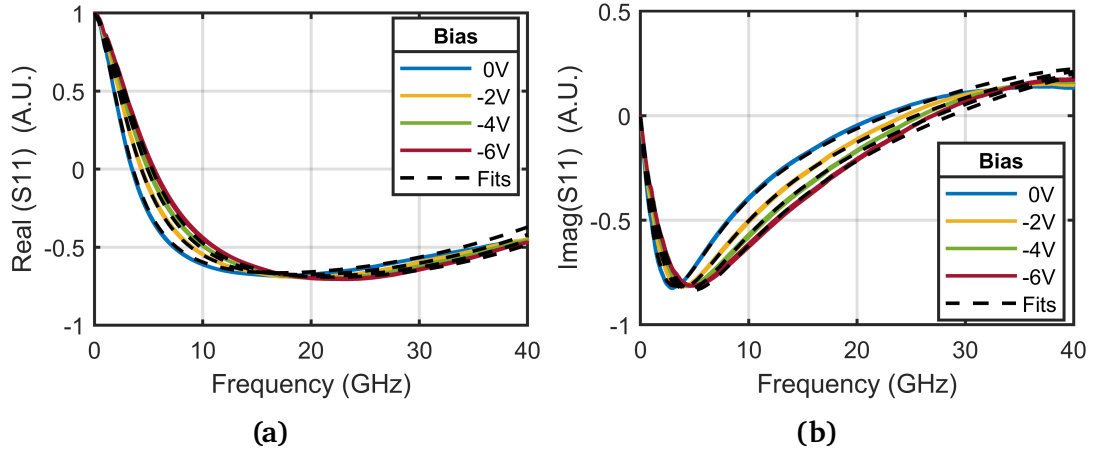


Figure 4.20: Fit of the equivalent circuit model to the measured S11 parameter at various bias voltages in terms of (a) its real part and (b) imaginary part.

The real and imaginary parts of the measured S11 parameter and corresponding fitted curves are shown in Figure 4.18 for a selection of bias voltages. Close agreement is seen between the measured and fitted data—certainly sufficient to use the equivalent circuit model to identify key trends in behaviour.

4.5.2 Deriving the S21 Frequency Response of the Lumped MZM from the Equivalent Circuit Model

With an accurate equivalent circuit model determined for the lumped MZM, the load voltage V_{load} that falls across the junction capacitance C_j at various frequencies can now be determined. The transfer function that gives this voltage with respect to the source voltage is the S21 parameter. Due to the multiple circuit

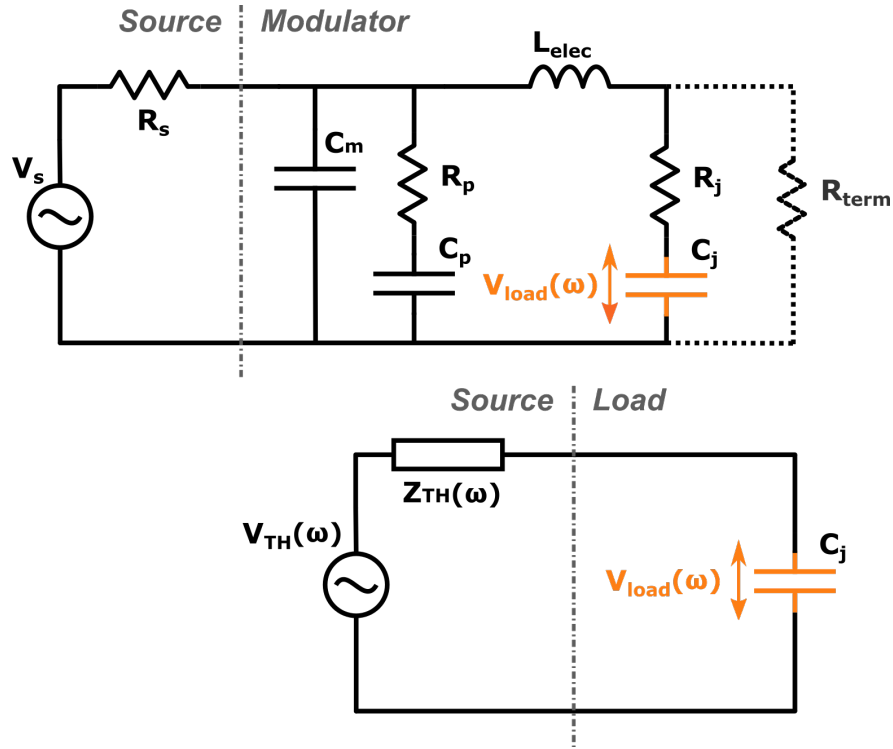


Figure 4.21: Conversion of the lumped MZM's equivalent circuit model from its standard form to its Thévenin equivalent.

elements present, it is convenient to use the Thévenin equivalent of the circuit in order to simplify the treatment. By doing so, the voltage source and various impedances for the lumped MZM can be replaced by an equivalent combination of a single voltage source V_{TH} in series with a single impedance Z_{TH} . Figure 4.21 shows the resulting Thévenin equivalent circuit transformation for the lumped MZM and voltage source. As later work in this chapter will analyse the effect of having a matching termination present in the circuit, an optional termination resistor R_{term} has been included in the treatment, shown in dotted lines. The full derivation of the Thévenin circuit is provided in Appendix C.

The Thévenin voltage is given by

$$V_{TH}(\omega) = V_s(\omega) \cdot \frac{Z_1}{Z_2} \cdot \frac{Z_{Rt}}{Z_{Lelec} + Z_{Rt}}, \quad (4.13)$$

where, for ease of writing, we have set $Z_1 = Z_{Cm} \parallel (Z_{Rp} + Z_{Cp}) \parallel (Z_{Lelec} + Z_{Rt})$ and $Z_2 = Z_{Rs} + Z_1$. The Thévenin impedance is given by

$$Z_{TH}(\omega) = \left(\left[Z_{Rs} \parallel Z_{Cm} \parallel (Z_{Rp} + Z_{Cp}) + Z_{Lelec} \right] \parallel Z_{Rt} \right) + Z_{Rj}. \quad (4.14)$$

The load voltage across the junction capacitance C_j is then given by

$$V_{load}(\omega) = V_{TH}(\omega) \cdot \frac{Z_{Cj}}{Z_{TH}(\omega) + Z_{Cj}}. \quad (4.15)$$

The crucial transfer function of the circuit that determines the MZM frequency response is then

$$H_{vload}(\omega) = \frac{V_{load}(\omega)}{V_s(\omega)}. \quad (4.16)$$

The frequency response as modelled by this equation will be compared with the measured frequency response of the lumped MZM later in Figure 4.23 after one more frequency-dependent effect is accounted for.

4.5.3 Effect of Optical Transit Time on Lumped MZM Frequency Response

Once the voltage transferred to the capacitor C_j has been determined from Equation 4.16, the final frequency-dependent effect that must be considered at high frequencies is the finite transit time of the light from one end of the phase shifter to the other. Here, the derivation of this effect follows the treatment in [64]. For electrodes below the lumped element cut-off length, the voltage present in an MZM arm at a given instant in time can be assumed to be identical regardless of the position on the electrode. Optical phase shift occurs as the light propagating in the MZM arm sees a change in effective index due to applied voltage. This effective index change is then a function of voltage, which is in turn a function of time, i.e., $\Delta n_{eff}(V(t))$.

The transit time of the light through the phase shifter is not an infinitesimal quantity. Rather, it can be expressed as

$$\tau_g = \frac{L_{mzm}}{v_g}, \quad (4.17)$$

where v_g is the group velocity of the light, given by $v_g = c/n_g$. Here, c is the speed of light in vacuum and n_g is the optical group index of the silicon waveguide obtained from Equation 2.5. Crucially, τ_g can be significant with respect to the time-variation of the voltage. As a result, the instantaneous voltage present along the electrode at the time the light enters the phase shifter will have changed in value by the time light exits the phase shifter. This weakens the overall modulation effect, as the light essentially sees an averaged or “smeared-out”

section of the drive waveform.

To see the impact of this effect on the frequency response, we can consider a single-toned, sinusoidal voltage of amplitude V_0 and angular frequency ω which is applied to the phase shifter. It can be assumed to be constant at any point z along the phase shifter at a given moment in time. Therefore, the voltage the light sees is simply given by $V = V_0 e^{i\omega t} = V_0 e^{i\omega(t_0 + z/v_g)}$. Here, t_0 is the time at which the light enters the phase shifter, and can be set to 0 without loss of generality for the purposes of this analysis.

Adapting this into the expression for optical phase shift due to a refractive index change (Equation 2.4), the cumulative phase shift experienced by the optical wave over the length of the phase shifter is then given by

$$\begin{aligned}\Delta\phi &= \frac{2\pi}{\lambda} \int_0^L \Delta n_{\text{eff}}(V_0 e^{i\omega z/v_g}) dz \\ &= \frac{2\pi}{\lambda} v_g \int_0^{\tau_g} \Delta n_{\text{eff}}(V_0 e^{i\omega t}) dt,\end{aligned}\tag{4.18}$$

where the second integral is reached through a change of variable: $z = v_g t$. In the small-signal regime, Δn_{eff} can be assumed to be linear with voltage, and so can be taken outside the integral, leaving only the exponential term inside. After carrying out the integration, substituting $v_g = L_{\text{mzm}}/\tau_g$, and normalising to low frequency, the frequency response as a result of the optical transit time can be obtained as

$$H_{op}(i\omega) = \frac{1 - \exp(-i\omega\tau_g)}{i\omega}.\tag{4.19}$$

The magnitude of this frequency response simplifies nicely to a sinc function:

$$|H_{op}(i\omega)| = \frac{\sin \tau_g \omega}{\tau_g \omega} = \text{sinc}(\tau_g \omega/2).\tag{4.20}$$

From this, we can obtain a 3 dB frequency cut-off for a given MZM length due the optical transit time as

$$f_\tau \simeq \frac{0.44}{\tau_g} = \frac{0.44c}{L_{\text{mzm}}n_g}.\tag{4.21}$$

L_{mzm}	f_τ
(mm)	(GHz)
0.5	68.8
1.0	34.4
1.5	22.9
2.0	17.2

Table 4.2: 3 dB frequency cut-off due to optical transit time effect for $n_g = 3.84$.

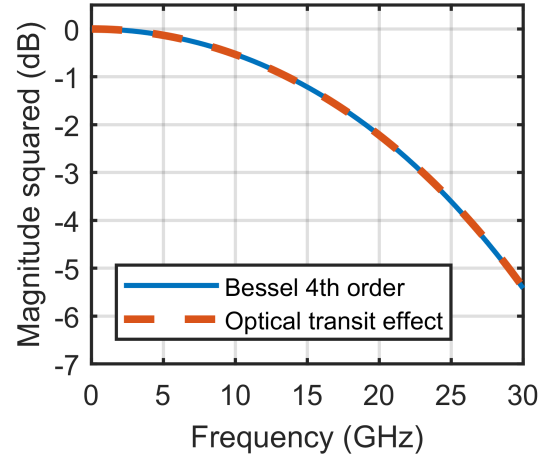


Figure 4.22: Similarity of optical transit time effect and fourth-order Bessel response. $L_{mzm} = 1.5$ mm.

Table 4.2 shows this value of f_τ for various lengths of MZM, where the value used for n_g is 3.84, as calculated in Section 2.2.2. As could be expected, for shorter MZMs of less than 1 mm length, the 3 dB bandwidth caused by this effect alone is high. For the 1.5 mm lumped MZM studied in this work, however, the impact of an f_τ of 22.9 GHz should be investigated for the 25 Gbaud PAM4 signals in question. As a judge of its potential impact, Figure 4.22 plots the magnitude of this frequency response as given by Equation 4.20. (The group delays of both responses are flat, and are not plotted). Notably, the roll-off is virtually identical over this range of frequencies to that of a fourth-order Bessel filter with the same 3 dB bandwidth. Given that the 3 dB bandwidth in question is 91.6 % of the 25 Gbaud symbol rate, such a roll-off can be expected to have minimal impact on the signal quality.

Now that a suitable expression has been determined for estimating the effect of the optical transit time of the light through the MZM, the measured and modelled frequency responses can be compared. Figure 4.23 plots the measured frequency response of the lumped MZM alongside the frequency response as determined from its equivalent circuit model (Equation 4.16) in the cases where the model does and does not include the additional effect of the optical transit time. On first viewing, the modelled and measured frequency responses appear to differ significantly; the measured response has a 3 dB bandwidth of 1.9 GHz while the corresponding value for both the modelled cases is 4 GHz. However, focussing on the the low frequencies from 0 GHz to 1 GHz, it can be seen that there is a steep initial drop in magnitude for the measured response that causes the majority of

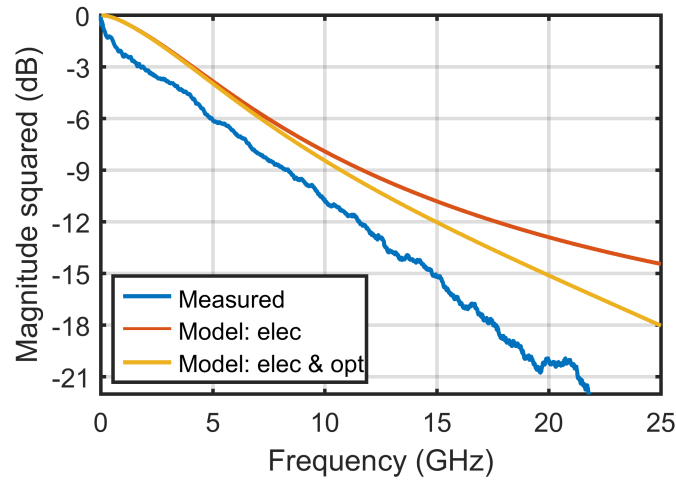


Figure 4.23: Comparison of the measured S21 frequency response of the lumped MZM with the model (with and without the optical transit time effect). The bias voltage is -3 V.

the difference in the curves. From this point onwards, the slope of the measured and modelled responses are quite similar. The electrical model (without the optical transit effect) follows the measured roll-off until approximately 10 GHz, while the addition of the optical transit effect to the model brings the measured and modelled responses closer to show a similar roll-off until 15 GHz.

The reason for the drop in bandwidth at low frequencies for the measured case is unclear, although a similar feature is observed in the measured frequency response provided by the device foundry for a shorter lumped 0.5 mm MZM [139]. When checking the literature, a very similar lumped MZM of the same length as our device does not show such a feature, instead presenting a flat response up to 1 GHz before beginning to roll off (Fig. 10 of [101]). Because of this, we assume that the measured drop in bandwidth at very low frequencies is particular to devices in this wafer run, rather than a universal characteristic of lumped MZMs, and so can be ignored in our general model of the frequency response.

As discussed in Section 2.4, 1.5 mm MZM electrodes can reasonably be considered as lumped elements up to 10 GHz or 25 GHz, depending on the choice of the constant Δ in Equation 2.24. Notably, in Figure 4.23, the modelled and measured responses begin to diverge further in the region beyond 15 GHz. This can be attributed to the onset of travelling-wave effects over the length of the MZM for which the lumped model cannot entirely account. For example, the RF attenuation in the electrodes of a silicon photonic MZM is actually

determined by two effects: the conductor loss (or skin effect) which has a square-root frequency dependence, and the silicon dielectric loss from the PN junction which has a quadratic frequency dependence [58]. These effects become more noticeable as the length of the MZM increases. Hence, a more involved model of the MZM frequency response that looks at these travelling wave effects could take this into account. Such a model typically involves a longer simulation flow and knowledge of the electrode characteristics that are not immediately accessible, and then integrating the RF waves (both forward and reflected) and optical waves over the length of the MZM, taking their values in time and position into account [101, 80].

However, this difference at higher frequencies does not invalidate our derived frequency response model for lumped MZMs of this length. If needed, to further ensure lumped electrical operation in the upper region of frequencies, without loss of generality, the MZM could be implemented with a meandered design [83, 103, 104]. This has been shown to keep the given RF distances electrically short (around 0.25 mm) while maintaining relatively long effective MZM phase shifter lengths (around 1.5 mm). Such a design works by taking advantage of the tight bend radii that are possible with silicon waveguides, winding the phase shifter back parallel to itself multiple times in a hairpin fashion. This shortens the distances between the RF vias and the contact point of the driver on the electrode, minimising to an extent the length of the path for electrical signals to reach any point along the PN junction. As such, the analysis carried out using the lumped frequency response model in the rest of this chapter can be seen as a way of identifying key trends in lumped MZM performance for various driver parameters, while using a simple, convenient model.

4.5.4 Effect of Removing the Termination Resistor

With a model determined for the frequency response of the lumped MZM, various drive circuit conditions can now be conveniently analysed. The first analysis of this kind will start by examining the effect of keeping or removing the 50 Ω termination resistor on a lumped MZM. In this section, the terms *terminated* and *unterminated* will be used respectively for these conditions. (We no longer make a distinction between an unterminated TW-MZM and an unterminated lumped MZM, as was done previously in Section 4.2.) Additionally, as the circuit parameters derived from the S11 fitting in Section 4.5.1 showed a very high junction capacitance (attributed to high doping levels), an MZM with lower-

doped PN junctions ($5 \times 10^{17} \text{ cm}^{-3}$ for both p and n sides) will be used in the following simulations, with the junction capacitance C_j adjusted accordingly. The fitted junction resistance R_j was higher than the values of approximately 4Ω to 8Ω obtained using the modelling shown previously in Figure 3.16(b) and Figure 3.17(a) over doping levels of $5 \times 10^{17} \text{ cm}^{-3}$ to 10^{18} cm^{-3} , which is likely due to extra parasitic effects. Its value was therefore left unchanged from the fit. The other elements of the circuit are also kept fixed. Hence, the final circuit parameters serve as a typical, representative example of a more moderately doped SiP depletion-mode phase shifter with higher bandwidth than our measured device, while using the additional knowledge of the parasitic elements gained from the S11 fitting. The values used are therefore as follows: $V_{bias} = -1 \text{ V}$, $R_j = 10.6 \Omega$, $C_j = 350 \text{ fF}$, $L_{elec} = 75 \text{ pH}$, $C_m = 100 \text{ fF}$, $R_p = 1000 \Omega$, $C_p = 200 \text{ fF}$. The value of the termination resistor, when used, is $R_t = 50 \Omega$. Otherwise, it is set to a very large impedance to remove its effect from the circuit.

Removing the termination resistor from the MZM has a noticeable effect on the frequency response, as shown in Figure 4.24. The most obvious change is an increase of 6 dB in the power of the modulation at low frequencies, which corresponds to twice the effective voltage V_{load} falling across the junction capacitance. This is beneficial in reducing the required modulation voltage for a given ER. However, similarly to the experimental results of Section 4.2, the bandwidth decreases significantly without the termination: the 50Ω terminated case has a 3 dB bandwidth of 11.4 GHz while the unterminated case reduces to 6.2 GHz. The group delay is also degraded, as shown in Figure 4.24(b), with the unterminated case reaching a maximum group delay of 33 ps as opposed to the 17 ps of the terminated case.

To get a qualitative sense of the effect of these frequency responses on signals, simulated PAM4 eye diagrams for the terminated and unterminated MZMs in a back-to-back link are shown in Figure 4.25 for 10 Gbaud and 25 Gbaud. For a fair comparison, the set drive voltage is adjusted to ensure that the same V_{load} falls across the junction capacitance in both the terminated and unterminated cases, i.e., the ER is always 3.5 dB, but the peak-to-peak voltage is 1 V in each arm for the unterminated MZM and 2 V in each arm for the terminated MZM. Comparing the signals at 10 Gbaud, both the terminated and unterminated cases can be considered as being good-quality, open eyes, with the unterminated MZM showing slightly greater eye closure than that of the terminated MZM. This can be expected, as 10 Gbaud is a relatively low-speed signal for both cases, given the MZM bandwidths in question. However, there is a marked difference in

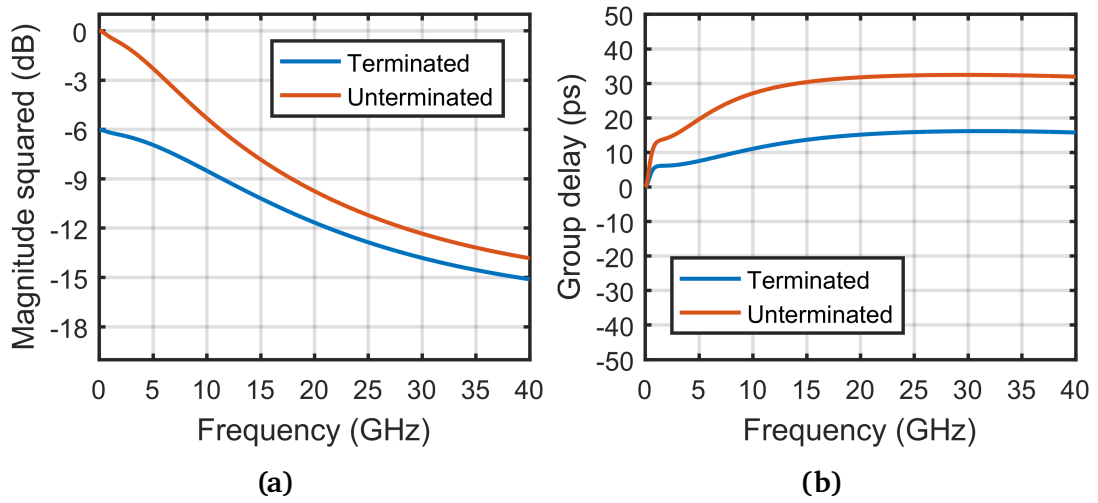


Figure 4.24: Comparison of the frequency response in terms of (a) amplitude squared and (b) group delay vs frequency for the LRC model with and without the $50\ \Omega$ termination.

eye quality when considering the 25 Gbaud signal. The eye diagram for the terminated MZM at 25 Gbaud is more impaired in comparison with its 10 Gbaud equivalent, but still exhibits a relatively open eye. The corresponding 25 Gbaud eye for the unterminated MZM shows complete eye closure, as could be expected for any system with ~ 6 GHz transmitter bandwidth.

This comparison is shown quantitatively in Figure 4.26, where the corresponding BER curves are shown against optical power at the receiver. Here, and for the rest of this chapter, only one FEC threshold is shown to de-clutter the graphs; a single value of 10^{-3} is used as a balance between HD-FEC and KP4-FEC. While 10 Gbaud PAM4 is not unduly stressed by the absence of the termination, 25 Gbaud PAM4 is less forgiving, with the unterminated MZM flooring at error rates well above the FEC threshold. Also shown on this graph are the equivalent BER curves obtained when the effects of the optical transit time on frequency response (Equation 4.19) are not included in the simulation. The resulting difference in the curves are very small for all cases due to relatively high bandwidth of this effect ($f_\tau = 22.9$ GHz), and so this effect will be omitted in future simulations in this chapter. This has the advantage of simplifying the analysis and discussion of the MZM to deal only with our equivalent circuit model.

From viewing Figures 4.25 and 4.26, clearly a solution must be found to improve the bandwidth if an unterminated, lumped MZM is to be used for 50 Gb/s PAM4.

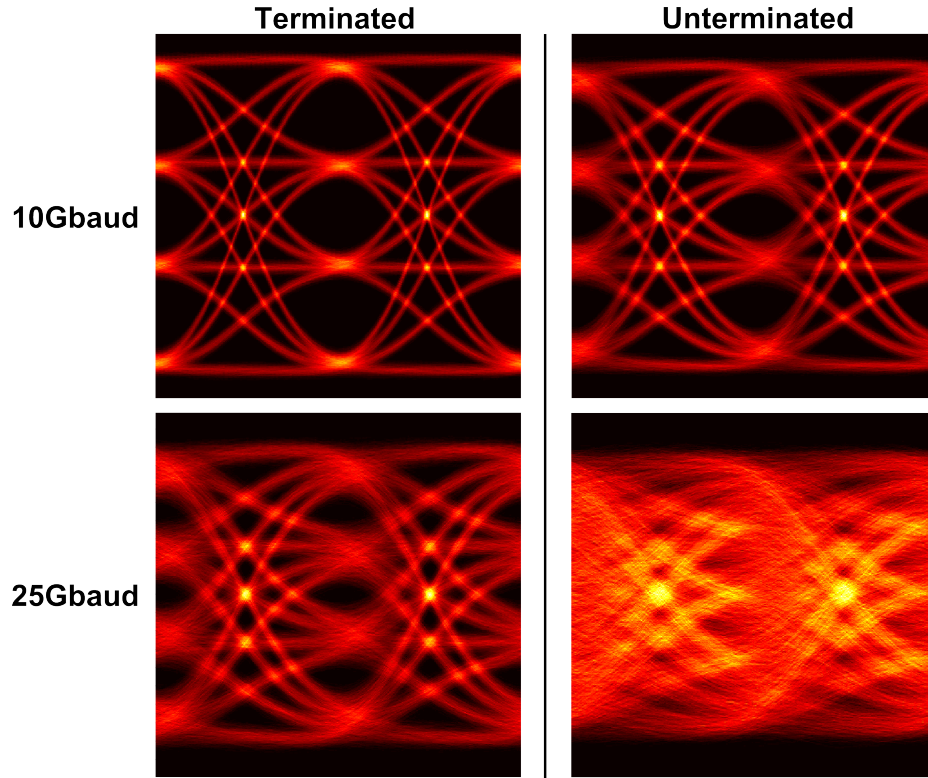


Figure 4.25: Simulated PAM4 eye diagrams at 10 Gbaud and 25 Gbaud for the terminated and unterminated MZM.

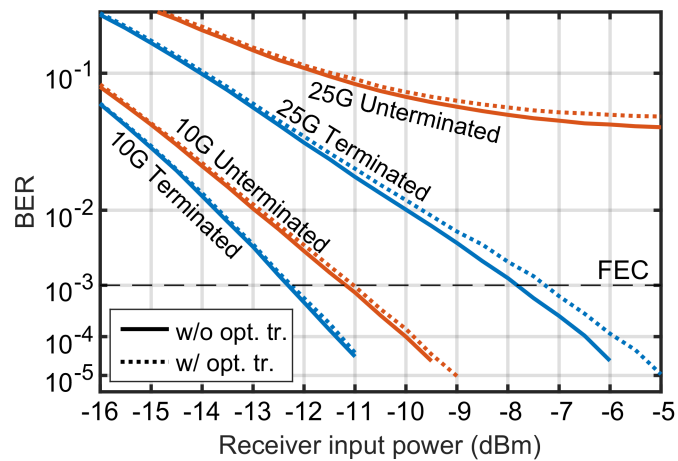


Figure 4.26: Simulated PAM4 BER curves at 10 Gbaud and 25 Gbaud for the terminated and unterminated MZM, with and without the effects of the optical transit time.

Methods of doing so will be analysed further in Section 4.6, but first, the energy consumption of the terminated and unterminated MZM will be discussed and compared.

As described in Section 4.3.1, the power consumed in the equivalent circuit can be calculated in the frequency domain by Equation 4.3 or in the time domain by Equation 4.4. To see how the power dissipated in the circuit varies with frequency, Figure 4.27(a) plots the power consumed for single-toned sinusoidal waveforms for both the terminated and unterminated MZM. For a fair comparison, i.e., to drop the same V_{load} across the junction capacitance at DC conditions in both cases, the peak-to-peak source voltage is doubled for the terminated MZM (2 V) with respect to the unterminated MZM (1 V). Similarly to the simple RC circuit in Figure 4.10(b), the power consumption of the unterminated MZM starts from zero at low frequencies, before increasing and flattening out as the resistive power starts to dominate at higher frequencies. In contrast, the terminated MZM starts with significant power consumption from DC due to the presence of the parallel termination resistor, before then increasing with frequency like the unterminated MZM. Naturally, the doubled source voltage required for the terminated MZM also contributes to its larger power consumption (by a factor of 4).

Figure 4.27(b) extends this analysis to a more meaningful metric for PRBS waveforms, plotting the energy consumed per bit for NRZ waveforms against symbol rate. Again, the peak-to-peak voltages in question are 2 V and 1 V, respectively, for the terminated and unterminated MZMs. The energy dissipated per bit is extremely high for the terminated case at low frequencies due to the 50 Ω termination resistor drawing a high current and dominating the energy consumption. As the frequency increases, the energy per bit decreases for the terminated MZM. This is effectively due to a more “efficient” use of the device, where the increase in symbol rate counteracts the high draw of power at lower frequencies, and wins over the relatively small increase in power dissipation in the higher frequencies. Contrastingly, the unterminated MZM exhibits a similar trend to that of the simple RC circuit in Figure 4.10(d), where the energy per bit increases slightly from low bit-rates before levelling out at approximately 250 fJ/bit, well before reaching the pJ/bit order. As indicated in the figure, the difference between the two cases is significant, with the terminated MZM consuming 12.9 times more energy per bit than the unterminated MZM at 10 Gbaud, and 6.1 times at 25 Gbaud.

The reason for this large difference in energy consumption can be seen clearly

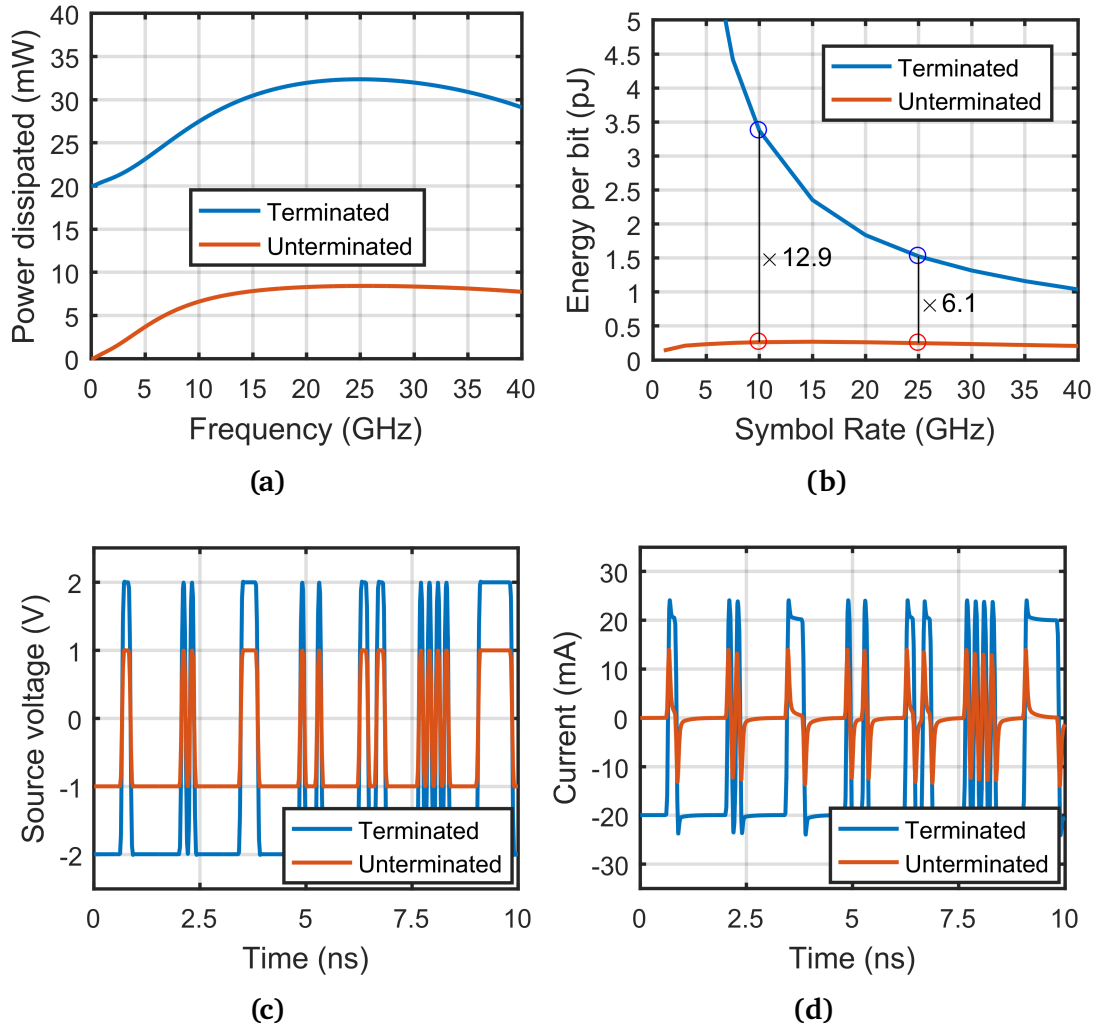


Figure 4.27: Comparison of the terminated and unterminated MZM in terms of (a) energy dissipated for sinusoidal waveforms and (b) energy dissipated per bit for NRZ waveforms. Also compared for the MZMs are the waveforms for (c) source voltage and (d) drawn current in the case of 10 Gbaud NRZ.

in the time domain in Figures 4.27(c) and 4.27(d), where the source voltage and drawn current waveforms are shown at 10 Gbaud for the terminated and unterminated cases. The source voltage is doubled in the terminated case to provide the same load voltage as the unterminated MZM. Comparing the current waveforms, it is clear that in the unterminated MZM current is only drawn during bit transitions due to its capacitive loading, while the terminated MZM keeps drawing current throughout the symbol period. Recalling that the power consumed depends on the dot product of the source voltage and current waveforms, we see that the terminated case will naturally be more power hungry. Indeed, for these drive conditions, the terminated case consumes 3.4 pJ/bit for

10 Gbaud NRZ, while the unterminated case consumes significantly less energy at 0.26 pJ/bit.

Although a comparison of the energy consumption for the unterminated and terminated MZM at 10 Gbaud is fair, given their relatively similar performance in Figure 4.26, the same cannot be said for a comparison at 25 Gbaud. The lack of termination leads to heavily impaired transmission at this speed, and a discussion of the energy consumption is not meaningful. Therefore, methods of improving a lumped MZM's bandwidth will be investigated in detail in the next section, with attention paid to the resulting increase in energy demanded in the equivalent circuit.

4.6 Improving Lumped MZM Bandwidth through Driver Considerations

4.6.1 Method 1: Use Predistortion

As described in Sections 4.3.3 and 4.4, a bandwidth limited system can be precompensated at the transmitter to produce an ideal signal at the receiver, provided the frequency response of the system is known and the sufficient drive voltage is available. In the following sections, it will be assumed that the driver can increase its peak-to-peak source voltage as necessary to boost the high frequencies of the signal. However, the resulting increase in energy dissipation and drive voltage will be noted. Figures 4.28(a) and 4.28(b) show respectively the magnitude and group delay of the frequency response of the modelled terminated MZM, unterminated MZM, and the unterminated MZM with precompensation designed to give a fourth-order Bessel response with 3 dB bandwidth of 25 GHz. As could be expected, the frequency response of the precompensated MZM is ideal, with a smooth roll-off in amplitude and zero group delay. The resulting simulated 25 Gbaud PAM4 eye diagrams of the three cases are shown in Figure 4.29. Both the terminated and unterminated MZM show bandwidth impairments, as seen before, but the precompensated unterminated MZM exhibits an ideal eye.

However, such precompensation does require additional drive voltage and, hence, energy consumption, if the ER is to kept the same before and after precompensation. The 25 Gbaud PAM4 signals in Figure 4.29 consume 0.42 pJ/bit for the terminated MZM, 0.07 pJ/bit for the unterminated MZM,

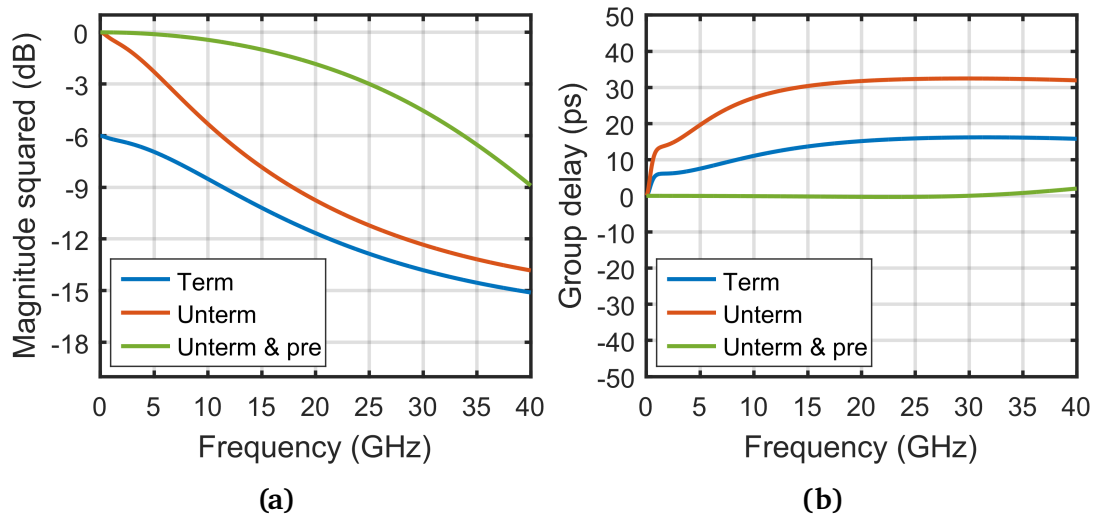


Figure 4.28: The effect of digital precompensation on the frequency response of the unterminated MZM for (a) amplitude and (b) group delay. The response of the terminated MZM is shown for comparison.

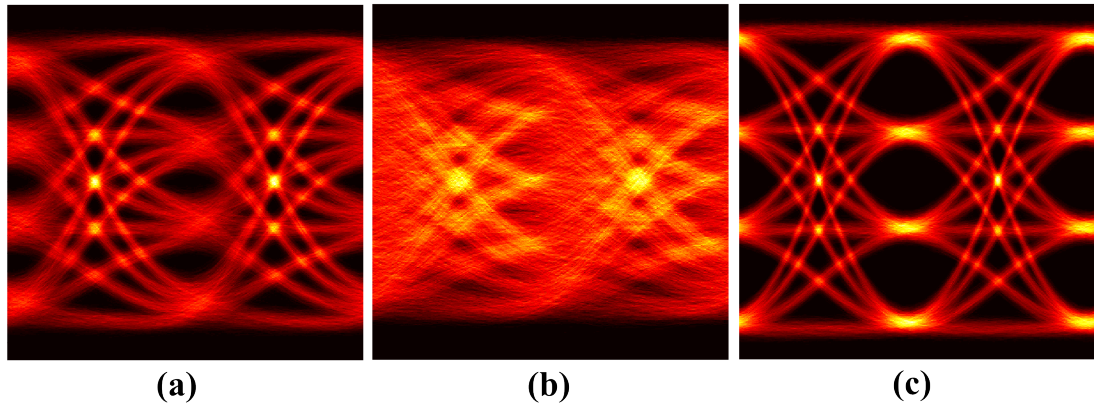


Figure 4.29: Comparison of the simulated 25 Gbaud PAM4 eye diagrams produced by the various frequency responses in Figure 4.28. The eyes are respectively from the (a) terminated MZM, (b) unterminated MZM, and (c) the unterminated MZM with precompensation.

and 0.19 pJ/bit for the precompensated unterminated MZM. To see how this increased energy cost comes about with respect to the frequency characteristics of the modulators, Figure 4.30 shows the energy dissipated per cycle for a frequency sweep of single-toned, sinusoidal drive voltages through the three MZMs, i.e., an extension of what was shown for the simple RC circuit in Figure 4.10(c). The peak-to-peak source voltage is 1 V for both the terminated and unterminated MZM. As the precompensated MZM has a matched filter applied (as described above) designed for ideal 25 Gbaud NRZ/PAM4 signals, the voltage of the input sinusoid driving it increases (from 1 V) according to the

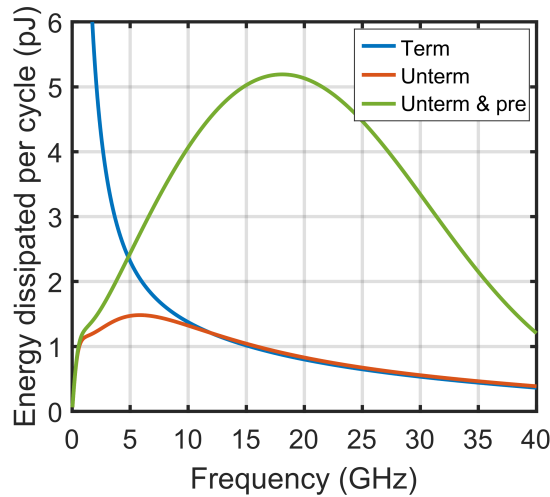


Figure 4.30: Energy dissipated per cycle for sinusoidal drive waveforms in the circuit in the case of the terminated MZM, unterminated MZM, and precompensated unterminated MZM.

magnitude of the filter at a given frequency. As a result, the precompensated case clearly consumes more energy per cycle as frequencies increase, before eventually rolling off like the other cases.

The reason for the increase in energy consumption can be seen in a familiar way in the time domain with PRBS waveforms; Figures 4.31(a) and 4.31(b) show the voltage and current waveforms for a 25 Gbaud NRZ signal for the three cases. Just as shown previously in Figure 4.27, the terminated MZM requires a 2 V peak-to-peak source voltage to produce a 1 V peak-to-peak load voltage, and draws current throughout each symbol duration, while the unterminated MZM needs half the source voltage and draws current only at symbol transitions. Most noticeably, the precompensated-unterminated MZM still draws current only at the symbol transitions, but the required peak-to-peak source voltage has increased to 4.3 V, similar to that of the terminated MZM. The spike of voltage present at the start of each symbol is typical of a drive voltage that is precompensated to account for low link bandwidth. It counteracts the slow rise and fall times that are present in the time-domain as a result of insufficient bandwidth, giving a “net” ideal waveform, as shown in Figure 4.31(c). Although the precompensated-unterminated MZM still dissipates less energy per bit than the terminated MZM, the similar, doubled source voltage required for each is not ideal. Hence, the next section will examine another method of dealing with the low-bandwidth of lumped MZMs.

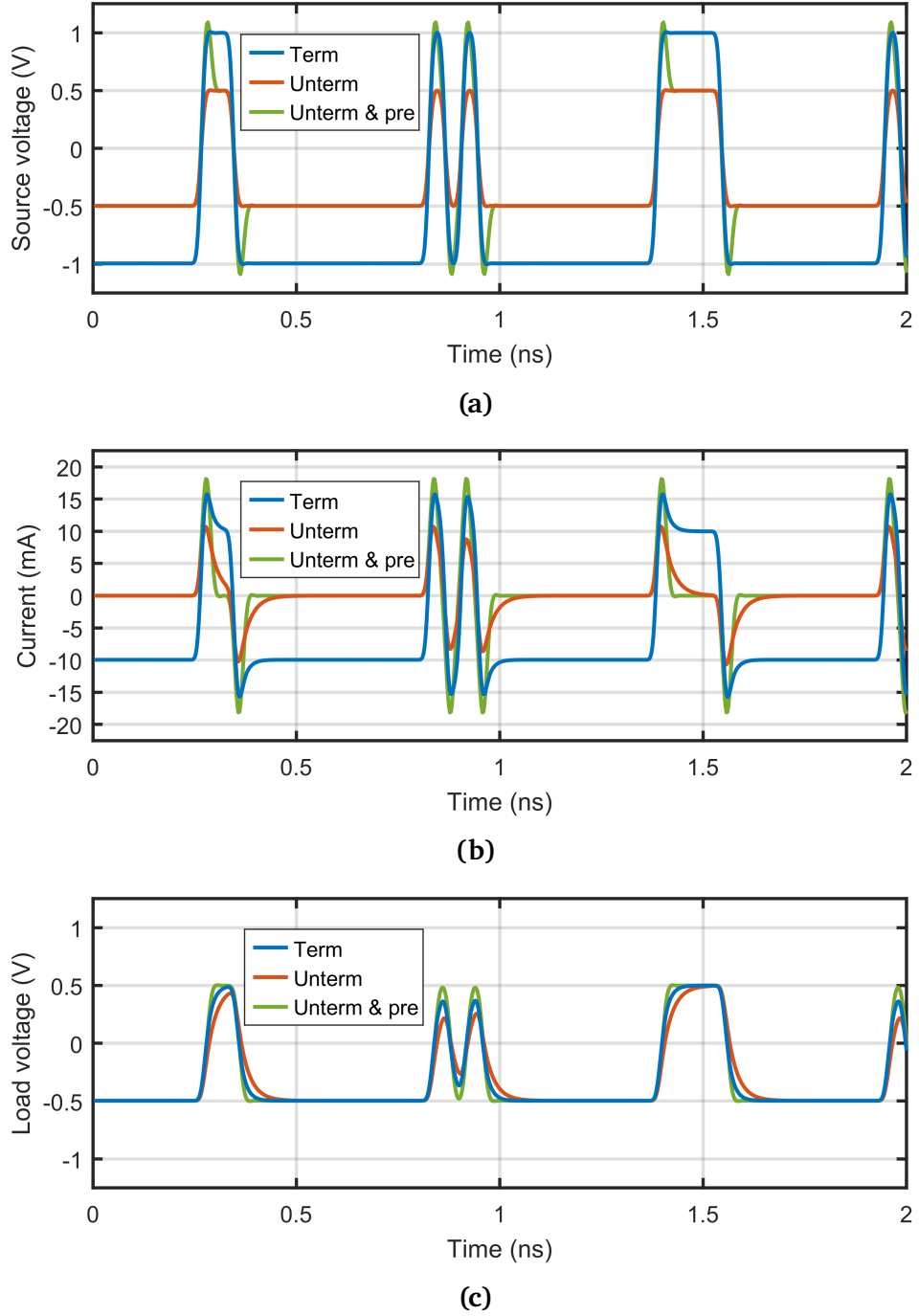


Figure 4.31: Comparison of the (a) source voltage, (b) drawn current and (c) load voltage waveforms for 25 Gbaud NRZ for the terminated, unterminated and precompensated-underminated MZMs.

4.6.2 Method 2: Reduce Source Resistance

Although precompensation solves the problem of the lumped MZM's low bandwidth, it is not an ideal solution for this modelled lumped MZM due to its doubling of the required peak-to-peak source voltage. Essentially, the lumped MZM simply has too low a bandwidth in its current form. As discussed in Section 3.4, the bandwidth of a lumped MZM can also be extended by reducing the source resistance from the typical $50\ \Omega$ to a much smaller value such as $5\ \Omega$. In this section, we will analyse the effect of reduced source resistance more thoroughly, now that a detailed equivalent circuit model for the lumped MZM has been derived.

For the switched-capacitor driver, the addition of a matching series capacitance halved the capacitance, and therefore boosted this bandwidth for the case of low source resistance further to bandwidths well beyond 25 GHz. However, given that there is a limited source voltage typically available from CMOS drivers, here we will not consider the use of the matching series capacitance, as this also doubles the source voltage requirements of the driver. The reduced boost in bandwidth as a result of this is not a major problem here. Recalling from Table 4.2 that the 3 dB bandwidth as a result of the optical transit alone is approximately 23.9 GHz, it is clear that the additional bandwidth provided by the series capacitors is essentially wasted, as the bandwidth is restricted by the optical transit time regardless. Of course, without the option of generating a four-level signal using the switched-capacitor bank, the PAM4 signal to be used must be generated digitally by the driver, through a DAC. The additional energy overhead of the DAC is not included in the following analysis; the energy dissipated is only that consumed by the circuit elements shown in Figure 4.16.

The effect of lowering the source resistance R_s on the magnitude and group delay of the lumped MZM's frequency response is shown in Figures 4.32(a) and 4.32(b). As R_s reduces, the magnitude rolls off more gently and the group delay decreases towards zero and flattens out. These two effects are, of course, conducive to a higher-quality modulated output waveform. Towards very low values of R_s ($5\ \Omega$, highlighted in red), the magnitude of the frequency response starts to rise above zero. This is caused by the inductance of the electrode (75 pH) taking effect and causing a small amount of inductive peaking.

The resulting 3 dB bandwidth is plotted against R_s for the lumped MZM in Figure 4.33(a). For comparison, the 3 dB bandwidth of an RC circuit is also shown (with resistance taken as the source resistance and the capacitance as

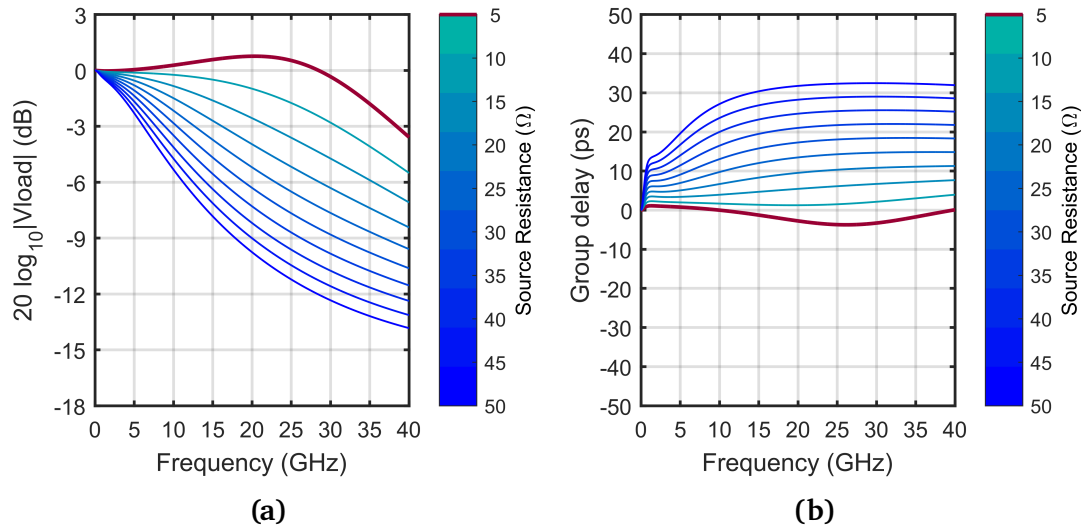


Figure 4.32: Effect of lowering source resistance on the frequency response of the unterminated MZM in terms of (a) magnitude and (b) group delay.

C_j). This demonstrates the effect of the parasitic components of the equivalent circuit model on the bandwidth. For example, they cause a 3 dB bandwidth reduction of approximately 4 GHz when the source resistance is 50 Ω and a 15 GHz reduction when the source resistance is reduced to 10 Ω . These parasitics, which will be present in any real circuit to a greater or lesser extent depending on implementation, are critical to include. For example, without them, the bandwidth of the equivalent circuit diverges to infinity as the source resistance is reduced further towards zero, which is certainly not achievable in practice.

The energy consumption of the lumped MZM circuit does vary when the source impedance is reduced, as shown in Figure 4.33(b), where the energy dissipated per bit for 25 Gbaud NRZ and PAM4 signals is plotted. The energy required increases slightly as the source resistance moves from 50 Ω to 20 Ω , before decreasing as the source resistance is lowered further. The presence of the parasitics in the circuit cause an additional energy consumption of 60 fJ/bit over the RC circuit for 25 Gbaud NRZ when the source resistance is 50 Ω , with a larger difference of 150 fJ/bit when the source resistance is reduced to 5 Ω . It should be noted that, for the same symbol rate and voltage swing, PAM4 is ~ 3.5 times more power efficient than NRZ in terms of the energy dissipated in the equivalent circuit. This is caused by the RMS values of voltage and current being lower by a factor of 0.75 in the PAM4 signal due to the intermediate symbol levels being of smaller amplitude than the outer symbol levels. With one in every two symbols on average in a PAM4 waveform being an intermediate level, the

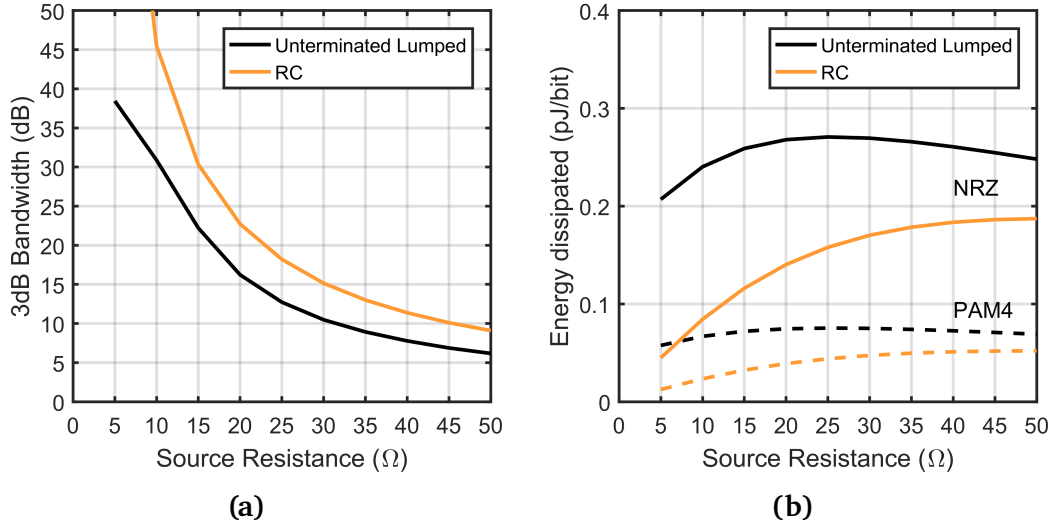


Figure 4.33: (a) Comparison of the 3 dB bandwidths as the source resistance is lowered for the LRC model and the simple RC model. (b) Comparison of the energy dissipated in the circuits for 25 Gbaud NRZ and PAM4 signals as the source resistance is lowered.

energy consumption is then reduced by a factor of $1/(0.75 \cdot 0.75 \cdot 2) \approx 3.56$.

Although designing the source resistance of the lumped MZM driver to be as small as possible does give the greatest 3 dB bandwidth, there are practical challenges in creating such a circuit; every electrical trace and transistor in the driver adds to the net source impedance, adding difficulty to the circuit design. Hence, the effect of lowered source resistance on the required optical power at the receiver for a given BER should be investigated in order to evaluate overall system performance and potentially find an optimum value for R_s .

Figure 4.34 shows the required Rx power for a BER of 10^{-3} for the lumped MZM as the source resistance is reduced from 50 Ω , for 25 Gbaud NRZ and PAM4 signals. As above, also shown for comparison is the result for the same RC circuit without parasitic elements. As could be expected, the required Rx power is high for a 50 Ω source resistance, standing at -3.6 dBm for the RC circuit in the case of PAM4, while the lumped MZM model is off the chart. For both cases, the required Rx power decreases with lowered source resistance, with both reaching a minimum of -10.5 dBm for PAM4 when R_s is 5 Ω , i.e., when the 3 dB bandwidth is maximised back in Figure 4.33(a). Near the minimum point, there is a 5 dB power penalty between the NRZ and PAM4 signals, which is to be expected due to the eye openings decreasing by a factor of three for the four-level signal. In addition, the required Rx power for the PAM4 signal rises

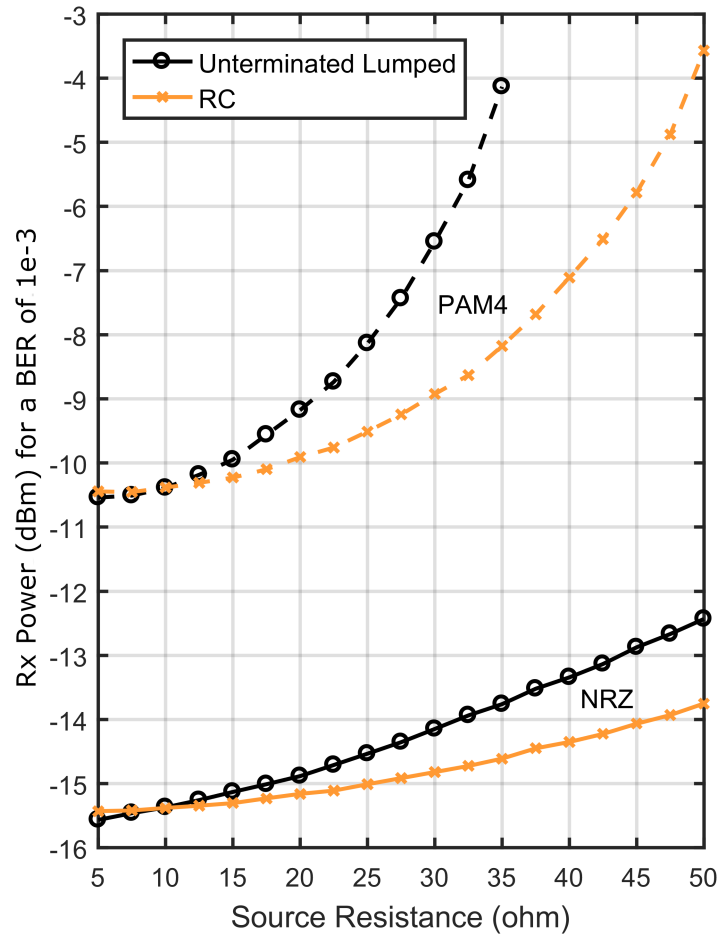


Figure 4.34: Required Rx power for a BER of $1e-3$ for 25 Gbaud NRZ and PAM4 signals when sent by the unterminated MZM. The equivalent for a simple RC circuit is shown for comparison.

more quickly with source impedance than that of the NRZ signal, due to PAM4's reduced tolerance for eye impairments. These results show that, for the best sensitivity, the source resistance of the lumped MZM's driver should be reduced as much as possible, although this design logic does give diminishing returns. For example, with a 25 Gbaud PAM4 signal, the requirements on low source impedance could be relaxed from $5\ \Omega$ to $25\ \Omega$, provided that a 2.5 dB penalty in required Rx power is acceptable for the system.

Comparing the behaviour of the RC model and the lumped MZM model in Figures 4.33(a), 4.33(b) and 4.34, it is clear that, despite the presence of its additional circuit elements, the lumped MZM essentially follows RC behaviour; the parasitics do impact the bandwidth and increase the energy consumption over that of the RC circuit, but the general trends are the same. This is largely to be expected, given that the only non-resistive or capacitive element in the circuit

is the inductance of the electrode L_{elec} . The fitted value for this (75 pH) can be considered to be quite low, given that such an exercise for a similar lumped MZM in another work [101] gave a much higher value of 375 pH. A potential reason for this difference is the geometry of MZM electrodes; modelling them as two parallel busbars, or long rectangular cuboids, the effective inductance associated with each is approximated by [140]

$$L = \frac{\mu_0 l}{2\pi} \left[\ln \left(\frac{d}{w+t} \right) + \frac{3}{2} \right], \quad w, d \ll l, \quad (4.22)$$

where l is the length of the electrodes, d is the separation of the electrodes as measured between two corresponding points on each, w is the electrode width and t is the electrode thickness. This equation holds predominantly for low frequencies, but can still be used to identify trends in how electrode inductance changes with geometry. Clearly, a larger inductance can be arrived upon for the same length MZM if greater separation occurs between the signal and ground electrodes, or if the electrodes are made thinner in height or less wide.

The exact trade-offs in the choice of MZM electrode geometry are beyond the scope of this thesis and will not be discussed further. However, given that MZMs can have larger values of electrode inductance than our fitted value (as evidenced by [101]), the resulting effect on the bandwidth and system performance is worth analysing. In addition, regardless of the inductance of the electrodes themselves, if wire bonds (as opposed to bump bonds) are used to electrically couple a driver to the MZM, significant inductance can be added. For example, using the rule of thumb of 1 nH/mm for a typical wire bond of diameter 25 μm [141], the larger inductance value of 375 nH in [101] can be reached from our fitted value of 75 nH with a short wire bond of length 300 μm .

As there are many possible values of L_{elec} depending on design, the following analysis will use the two aforementioned inductance values (75 pH and 375 pH) as representative examples of *low-inductance* and *high-inductance* cases, respectively. Figures 4.35(a) and 4.35(b) shown the frequency response of the circuit in terms of magnitude and group delay when the inductance has been increased to 375 pH. The effect of the inductance is now far stronger as the source resistance is lowered, with noticeable inductive peaking being exhibited below 25 Ω . The resonant frequency that determines the location of this peak occurs near $1/(2\pi\sqrt{LC})$, where C is the net value of the significant capacitances of the circuit and L is the inductance. Taking $L = L_{elec}$ and summing the parallel capacitances to get $C = C_j + C_m$ (the other capacitance C_p can be ignored

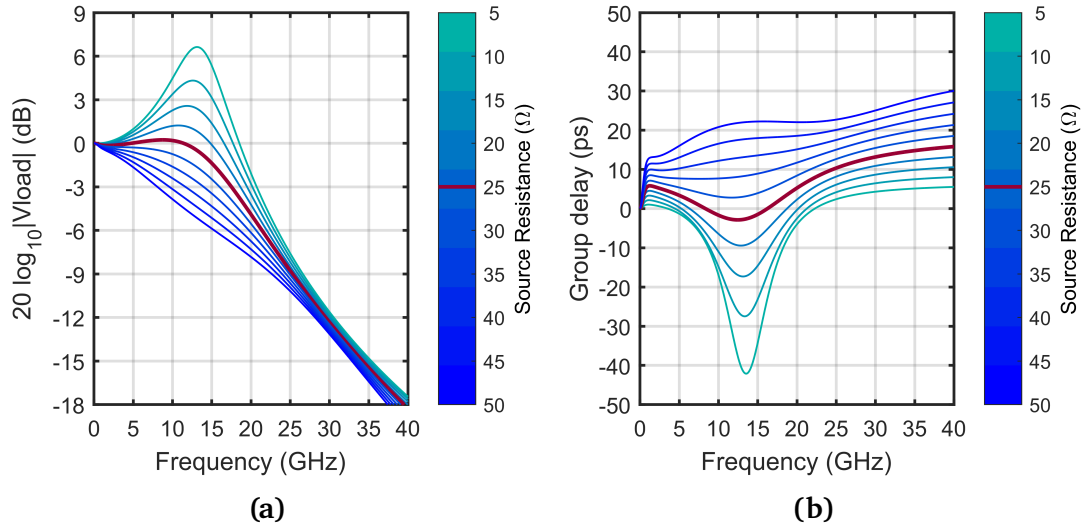


Figure 4.35: Effect of lowering source resistance on the frequency response of the unterminated MZM with high inductance in terms of (a) magnitude and (b) group delay.

due its very large corresponding series resistance R_s), the calculated resonant frequency is then 12.25 GHz, closely matching what is seen in the figures.

Such peaking naturally improves the MZM's 3 dB bandwidth, as demonstrated in Figure 4.36(a), where the bandwidths of the low-inductance and high-inductance cases are compared. The high-inductance MZM gives greater 3 dB bandwidth than the low-inductance MZM for source resistances from 50 Ω to 17.5 Ω , after which point inductive peaking causes the bandwidth to drop off quickly. There is a price to pay with this extra inductance, however. The added inductance causes greater energy consumption in the circuit, as shown in Figure 4.36(b). The increase is small (25 fJ/bit for NRZ) when the source resistance is a standard 50 Ω , but grows more significant as the source resistance is decreased, being 60 fJ/bit more energy intensive when the source resistance is 25 Ω , and requiring 200 fJ/bit more for a 5 Ω source resistance.

Although Figure 4.36(a) again shows that the maximum 3 dB bandwidth occurs at the lowest possible source resistance, the presence of strong inductive peaking means that such a design will not give the best system performance. Generally, the optimal frequency response for such a system is one that remains flat for longest (or rolls off slightly) towards the symbol rate, i.e., when the source resistance is 25 Ω in this case. To investigate this, Figure 4.37 shows the required Rx power against source resistance for the same simulation parameters as above, now comparing the system performance of the low-inductance MZM and the

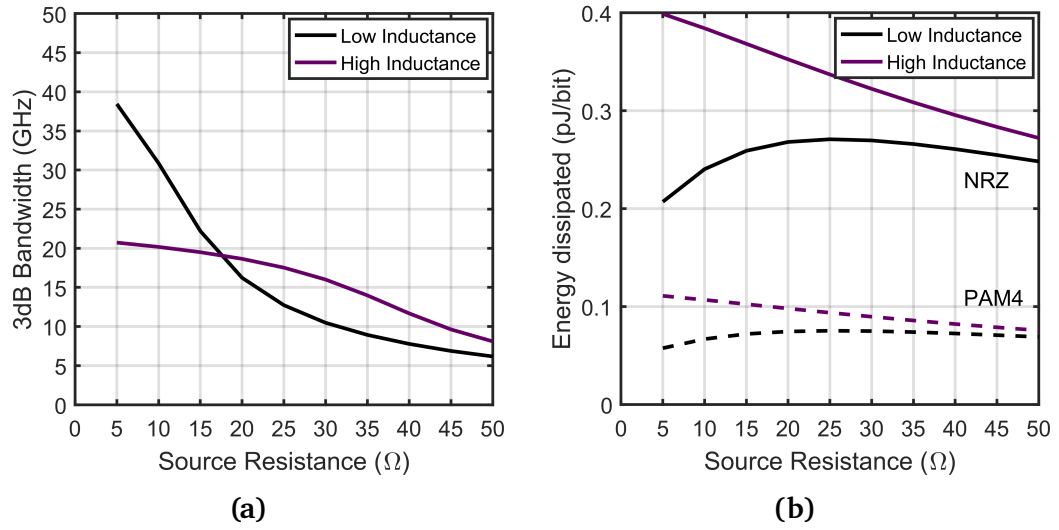


Figure 4.36: (a) Comparison of the 3 dB bandwidths as the source resistance is lowered for the lumped MZM with low inductance and high inductance. (b) Comparison of the energy dissipated in the circuits as the source resistance is lowered.

high-inductance MZM. The graph verifies that the optimal source resistance is now 25 Ω rather than the lowest value. For values of R_s above this point, the dominant system impairment is low bandwidth, while below this, over peaking from the inductance starts to take effect. As indicated in the figure, the regions to the left and right of this optimal source resistance can be classified respectively as the *inductive peaking regime* and the *low-bandwidth regime*.

There is minimal difference (less than 0.25 dB) penalty in the minimum required Rx power between the low-inductance and high inductance case; this is important, as it shows that the requirements on lowering source resistance can be relaxed for a lumped MZM that has an appreciable amount of inductance. Furthermore, although the trough-like shape of the sensitivity curve for PAM4 in Figure 4.37 suggests that best performance requires exactly 25 Ω source resistance, the sensitivity degrades by only 1 dB over the range of 18 Ω to 34 Ω . Again, this relaxes the design constraints for an implementation of such a driver.

As could be expected, 25 Gbaud NRZ is less sensitive to variations in source resistance than its 25 Gbaud PAM4 counterpart due to its simpler eye shape; the region of over-peaking from inductance does not have noticeable effect until approximately 10 Ω source resistance. In fact, the optimal source resistance for NRZ (20 Ω) is marginally lower than that of PAM4, as slight over-peaking actually benefits the simpler two-level signal.

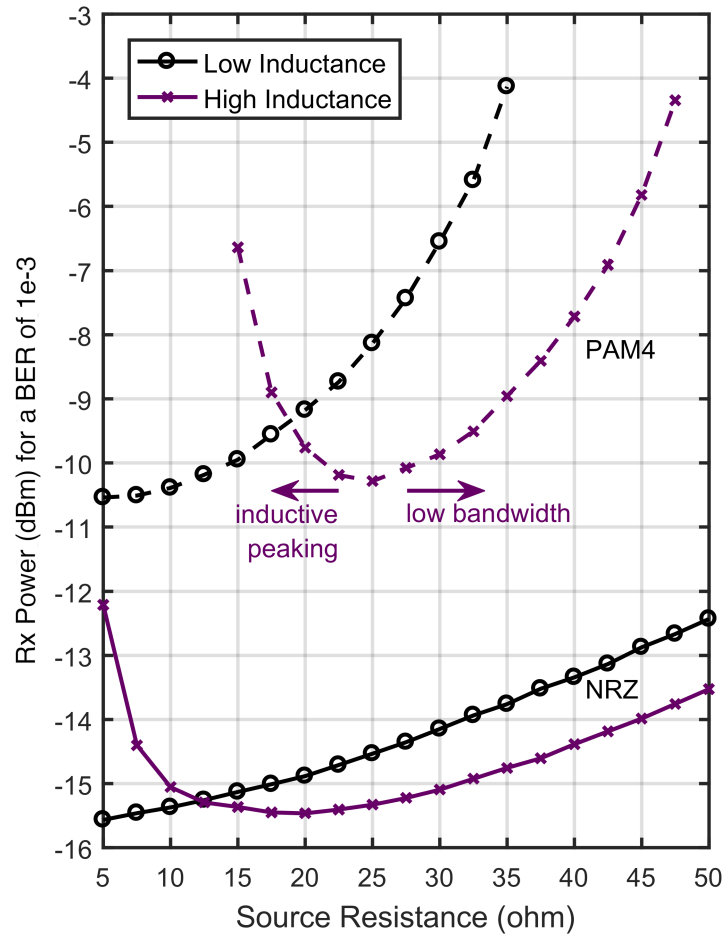


Figure 4.37: Required Rx power for a BER of $1e-3$ for 25 Gbaud NRZ and PAM4 signals from the lumped MZM, plotted against source resistance, comparing the cases of low and high inductance.

It is clear from the above analysis that the optimal choice of source resistance for a given inductance are linked, and vice versa. It would be possible to sweep more than just two values of inductance and obtain a three-dimensional equivalent for Figure 4.37, with inductance on another axis. However, as the testbed simulation for each point in this figure requires a sweep of BER against receiver power, this would be a lengthy process and so the analysis was deliberately restricted to two representative inductance values. (The four curves in Figure 4.37 required approximately 24 hours of simulation time; expanding the parameter space to an equivalent number of inductance values would have required weeks).

4.6.3 Method 3: Predistortion & Reduced Source Resistance

With two distinct methods proposed for improving the bandwidth of lumped MZMs, a question regarding a third naturally arises: what happens if *both*

precompensation and a lowered source resistance are implemented?

Figure 4.38 shows the simulated required Rx power for the same high-inductance MZM, now with and without precompensation to optimise the signal. The precompensation applied is a matched filter that is again designed to give a fourth-order Bessel response with a 3 dB bandwidth of 25 GHz. As could be expected, precompensation results in the best possible required Rx power regardless of the source resistance; provided sufficient source voltage is available from the driver and the frequency response is well known, the linear impairments of the system can be compensated perfectly ³. Notably, the required Rx power of the non-precompensated case at optimal source impedance matches that of the precompensated case. Certainly, any real implementation of an MZM and driver will have features in the frequency response that reduced source impedance and inductive peaking cannot perfectly compensate. However, this result does point towards the potential for a well-designed driver without precompensation to approach the performance of a precompensated scenario.

Although precompensation resolves the issue of bandwidth regardless of source resistance, the required source voltage and resulting energy dissipated are heavily affected by the type of bandwidth impairment in question; compensating for the effects of low bandwidth is not the same as compensating for the effects of over peaking. Figure 4.39(a) shows the required source voltage for precompensating the signal, plotted against source resistance. The non-precompensated case is shown for reference. The peak-to-peak source voltage is highest for the 50 Ω source resistance, with 1.8 V required in each arm. This is slightly less than the 2 V that was needed to precompensate the low-inductance MZM back in Figure 4.31(a), as the bandwidth is slightly better with higher inductance, even at a standard 50 Ω source impedance. The required source voltage decreases with the source resistance, until a knee is reached when $R_s = 22.5 \Omega$, very close to the source resistance for optimum receiver sensitivity. Beyond this point, the voltage needed remains relatively constant at approximately 1 V. Essentially, as the chosen source resistance strays into the low-bandwidth regime, the required drive voltage when using precompensation increases. In contrast, no extra drive voltage is needed in the inductive-peaking regime.

The resulting effect of precompensation on the energy consumption of the

³The marginally better sensitivity for the non-precompensated NRZ signal at 20 Ω source resistance is due to inductive peaking giving a (slightly) more optimal frequency response for this case than the chosen Bessel response produced by the matched filter.

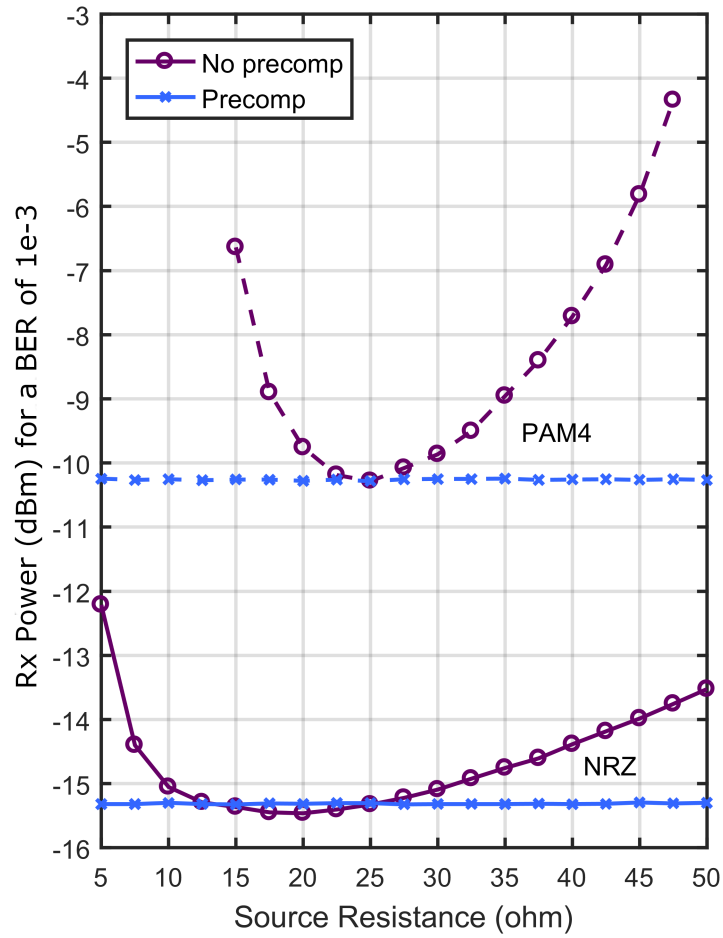


Figure 4.38: Required Rx power for a BER of $1e-3$ for 25 Gbaud NRZ and PAM4 signals when sent by the unterminated MZM without and with precompensation.

circuit is shown in Figure 4.39(b), where the energy dissipated per bit for 25 Gbaud NRZ and PAM4 waveforms is shown against source resistance. The non-precompensated case is again shown for comparison. The precompensated case dissipates approximately twice the energy of the non-precompensated case at standard $50\ \Omega$ source resistance. However, as established earlier, the bandwidth of the non-precompensated lumped MZM is too low at this source resistance for 25 Gbaud PAM4 transmission, so its energy saving is not a significant plus. When the system performance of both cases is equal (i.e., when $R_s = 25\ \Omega$), the energy dissipated is exactly equal. Below this point, in the inductive peaking regime, the precompensated case actually dissipates less energy per bit than the non-precompensated case.

The reason for this transition in energy performance can be seen clearly by looking at the precompensated waveforms for the drawn current and source voltage in the time domain. Figure 4.39(a) shows the drawn current for two

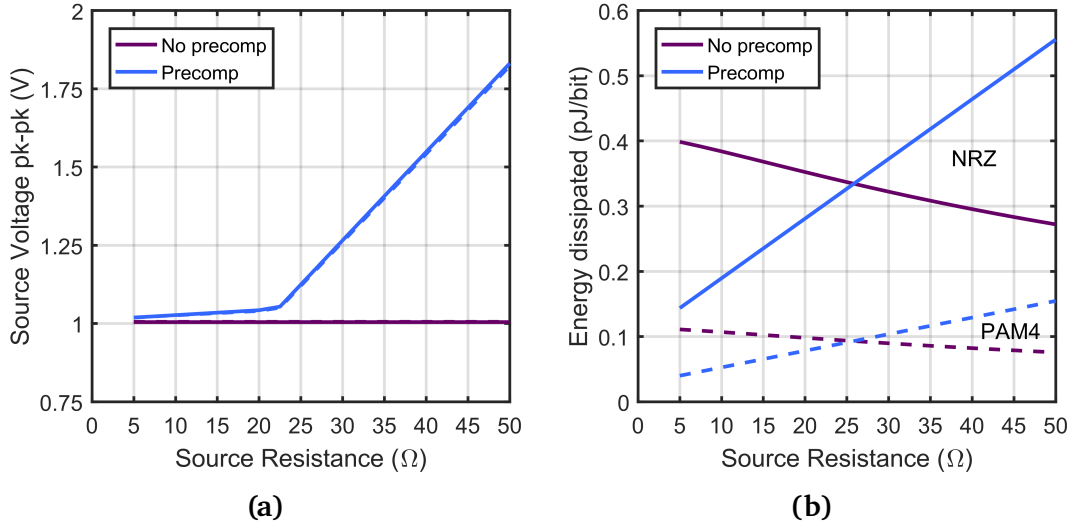


Figure 4.39: (a) Peak-to-peak voltage required for a 25 Gbaud NRZ signal with 4 dbER as source resistance is reduced for the unterminated MZM for the cases of without and with precompensation. (b) Corresponding energy dissipated in the circuit, shown for NRZ and PAM4.

example source resistances in the low-bandwidth ($R_s = 50 \Omega$) and inductive peaking regimes ($R_s = 20 \Omega$). For ease of analysis, 25 Gbaud NRZ is used rather than PAM4. Due to the precompensation, the current flowing in the circuit is identical for both cases. However, when viewing the corresponding source voltages in Figure 4.39(b), a clear difference is seen. In the poor bandwidth regime, naturally the peak-to-peak voltage has increased, providing a spike of voltage at the symbol transitions to account for the insufficient rise time. In the inductive peaking regime, however, the opposite occurs. The presence of the resonant peak in the frequency response means that certain frequencies in the signal now have too much power, and the source voltage must be reduced just after the symbol transitions in order to compensate and flatten the response. The power, i.e., the dot product of the source voltage and current waveforms, is then reduced for the precompensated, inductively peaked MZM. As a result, one potential for a low-power, low-voltage design of lumped MZM driver is to reduce the source impedance as much as possible, while adding inductive peaking and precompensation.

4.6.4 Summary of Bandwidth-Boosting Drive Methods

The drive options discussed in this chapter to increase the bandwidth of a lumped MZM are summarised in Figure 4.41, where the required source voltage

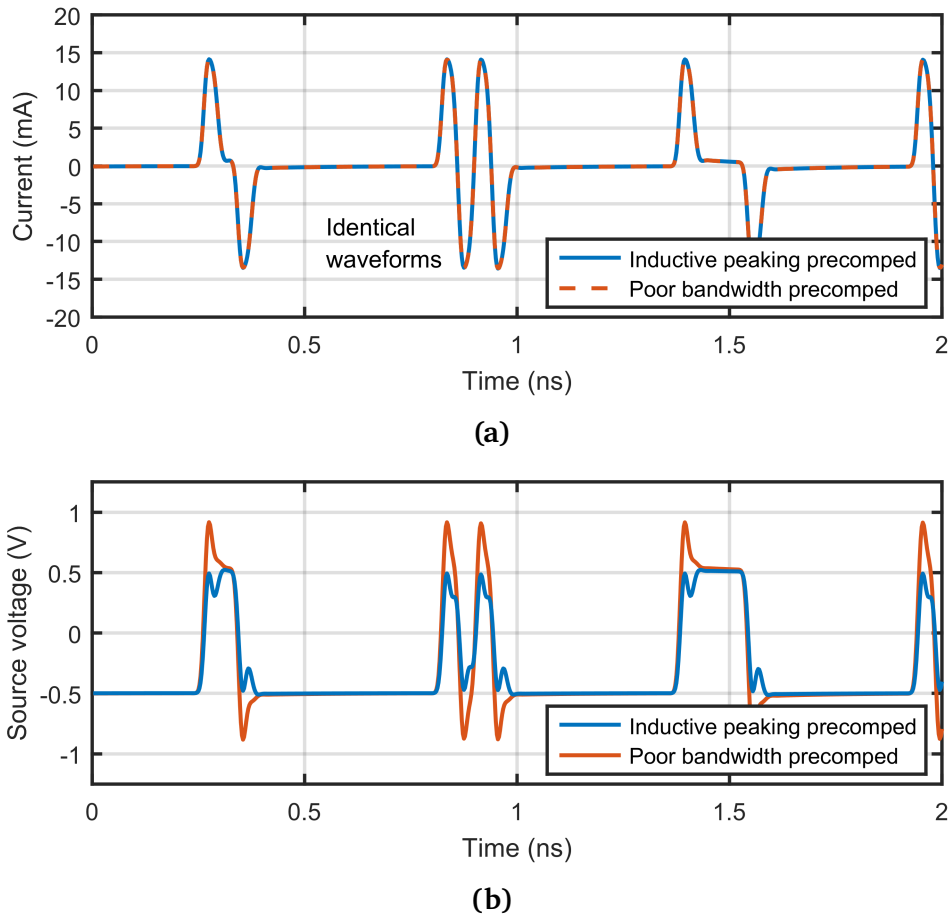


Figure 4.40: (a) Current drawn for unterminated MZM with precompensation in the cases of $50\ \Omega$ (poor bandwidth regime) and $20\ \Omega$ (inductive peaking regime) source resistance. (b) The precompensated source voltages for the two cases, showing how the poor bandwidth regime requires more voltage.

and energy dissipation per bit to achieve a 25 Gbaud signal with 4 dB ER are noted. The voltage V_{pp} in the figure refers to the peak-to-peak source voltage needed in each arm of the MZM when operating in push-pull fashion. For a $50\ \Omega$ drive scenario, using an unterminated, lumped MZM and applying precompensation causes the drive voltage requirement to increase significantly, exceeding that which would be required for a terminated MZM of the same length. If the constraint of a $50\ \Omega$ environment can be relaxed while still employing precompensation, lowering the source resistance until some inductive peaking occurs can offer the best combination of eye quality, power consumption and peak-to-peak voltage.

It should be emphasized that this work looks only at trends in energy consumption for various drive methodologies; the driver is modelled here as ideal voltage source in series with a simple source resistance and does not

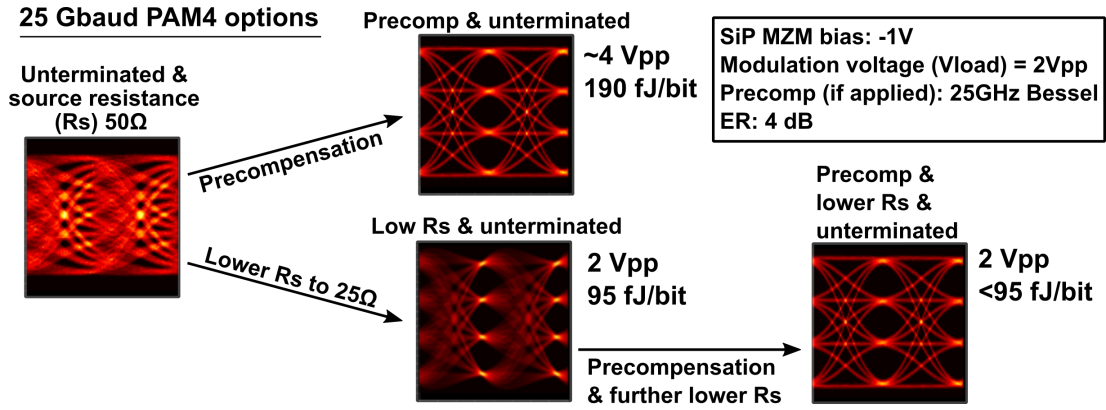


Figure 4.41: Summary of the discussed lumped MZM drive options for 25 Gbaud PAM4 in terms of required drive voltage to achieve a 4 dB ER, also showing the energy dissipated per bit and eye quality.

include any additional energy overheads such as the static power consumption of a DAC used to generate the precompensated electrical PAM4 signals. Instead, it focusses on the increased output swing for a given precompensation, which is crucial both in terms of the energy consumption and for the choice of silicon technology used (CMOS or BiCMOS). It should also be recognized that reducing the output impedance of the driver is not a trivial task and requires some trade-offs in design. For example, reducing the output impedance of a driver using current-mode logic (CML) will lead to an increase in power consumption in the driver as more current flows across the output/source resistors [142], making it a non-ideal solution for this case. A less energy-hungry approach is to drive the MZM with the output of a CMOS inverter, which naturally has a low source impedance and benefits from a small footprint [106, 102, 143]. Unfortunately, the binary nature of inverter outputs means that electrical precompensation of the signal is not possible with such a simple drive scheme. In this case, predistortion techniques would be restricted to inductive peaking such as presented here or alternative passive RC equalisation methods [106]. If electrical precompensation is desired, a linear driver amplifier would be required; such a driver with a relatively low source impedance (30Ω) has been demonstrated previously in the literature for an unterminated MZM [144].

4.7 Summary

This chapter focusses on the use of precompensation to boost the bandwidth of lumped SiP MZMs. In Section 4.2, the effectiveness of precompensation for lumped MZMs is demonstrated, where a lumped MZM and an unterminated TW-MZM with low electro-optic bandwidths are both shown to operate at 25 Gb/s with NRZ, PAM4 and EDB modulation formats after precompensation is applied. The addition of receiver-side equalisation then enables 50 Gb/s PAM4 and EDB operation over 10 km of fibre.

Section 4.3 focusses on the optimal design of precompensation filters, focussing first on RC circuits as a simple example of lumped MZM behaviour. Compensation of the magnitude alone of the RC frequency response is shown to improve system performance, but the impact of the phase of the frequency response is shown to cause residual sensitivity penalty, particularly for PAM4. This points to the importance of designing matched filters which compensate both the magnitude and phase of the response. The well-established theory for such filters is briefly presented.

In Section 4.4, the experimental setup used for the lumped MZM is revisited. A key limiting factor in the form of reflections in the RF cables at the transmitter side is identified and successfully compensated using long-length matched filters. This precompensation is then used to operate the lumped MZM at 50 Gb/s in PAM4 and EDB signal formats, without the need for receiver-side equalisation. Simulations are then performed to examine the precompensated lumped MZM's performance over fibre. EDB is shown to slightly outperform PAM modulation by <1.2 dB sensitivity for short lengths of fibre up to 5 km; overall, PAM4 provides slightly better reach of 12 km for the system in comparison with to the 8.5 km of EDB.

In order to look further into the energy dissipated in the lumped MZM, a suitable equivalent circuit model must be found. Section 4.5 focusses on this, starting with S11 measurements of a lumped MZM and then fitting a modified LRC model to the results. The circuit parameters obtained from this are then used to model the frequency response of the lumped MZM. The frequency-response effect of the optical transit time of light through the MZM is also modelled and determined to be reasonably small for the 25 Gbaud PAM4 signals of interest in this chapter.

In Section 4.6, using the equivalent circuit model obtained in the previous section and focussing on 25 Gbaud PAM4 and NRZ formats, various methods

of extending the bandwidth of lumped MZMs are examined. Particular attention is paid to the effect on required drive voltage and power consumption. Reducing the source resistance is shown to increase the bandwidth significantly, as would be expected from previous results in the thesis. The presence of reasonable amounts of induction in the circuit (from wire bonds, for example) is shown to result in the optimal source resistance moving from away from $0\ \Omega$ and towards a more moderate $25\ \Omega$, while achieving similar system performance and power consumption. This conveniently relaxes the design requirements on using very low impedance drivers. The use of precompensation is shown to increase the drive voltage (and thus, the power consumption) of the MZM significantly when used in a $50\ \Omega$ driver scenario. However, if precompensation is used in an inductive peaking regime (by reducing the source impedance), the required peak-to-peak drive voltage is lessened, and the overall power consumption is reduced.

Chapter 5

Conclusions

In a world that is becoming evermore interconnected, the future of data centres lies with low-power, high-density and high-speed optical transceivers. Silicon photonics offers such a solution through reliable, high-volume CMOS fabrication of such devices. The Mach-Zehnder modulator stands to be a key player in data centre links, with its ability to provide broad-spectrum transmission and high modulation bandwidths. In this thesis, analysis and optimisation of lumped silicon photonic MZMs for high-speed data centre links were presented using a combination of experiments and simulations.

5.1 Contributions Overview

Firstly, detailed characterisations are carried out of 1.5 mm dual-drive SiP lumped MZMs. Through these, the strong dependence of the phase-shift efficiency, absorption and modulation bandwidth of the silicon PN-junctions on bias voltage is identified. A detailed DC model of silicon phase shifters is then developed, which takes into account the optical mode distribution, doping concentrations and applied voltage to determine their phase-shift efficiency and absorption. This allows the V_π and insertion loss to be investigated for various MZM doping concentrations and bias conditions. A simple model for the RC bandwidth of the phase shifters is also developed, which uses the mobilities and distribution of the dopant elements to determine the junction resistance, and uses the dopant levels and bias voltage to determine the depletion width and, hence, the junction capacitance. With this model, the RC bandwidth of lumped MZMs is investigated for various doping profiles and lengths. The key limiting factor on the bandwidth is determined to be the 50 Ω source impedance of the RF drivers that are typically

used for these modulators.

Following this, drivers with low source impedance are investigated, focussing in particular on a novel, low-impedance, switched-capacitor PAM4 driver which enables multilevel modulation without the use of a typical electrical DAC. The DC phase shifter model developed earlier in the thesis is used to optimise the design and operation point of the lumped MZMs when used with the switched-capacitor driver. For the peak-to-peak voltages available from deep sub-micron CMOS processes (approximately 1 V), a balance must be found in the trade-off of signal ER and device insertion loss. Notably, relatively low doping levels of $2 \times 10^{17} \text{ cm}^{-3}$ are shown to give the optimal OMA for the switched-capacitor lumped MZM. Furthermore, provided that ample modulation bandwidth is available from the low-impedance driver, the optimal bias voltages for the PN junction phase shifters are found to be as low as possible; maximum OMA is achieved when the phase shifters are kept as close to their built-in voltage as the RF voltage swings allow.

Also investigated for the switched capacitor design is the effect on system performance of the silicon junctions' non-linear phase shift and variation of absorption with voltage. This is done firstly through LMR, and then by defining a suitable transmission link penalty as figure of merit. Through this, multilevel modulation formats are shown to be affected very little by the phase and absorption characteristics of the junctions, with essentially no penalty for PAM4 and less than 1 dB penalty for PAM8, even when employed with lumped MZMs of up to 2 mm in length and with peak-to-peak drive voltages of 2 V in each arm.

Next, the performance of the switched-capacitor design is evaluated in the context of a full silicon photonic link. The transmitter is a prototype switched-capacitor lumped MZM and the receiver is a silicon-germanium photodiode packaged with CMOS amplifier circuitry. 30 Gb/s PAM4 transmission is demonstrated over 10 km of fibre, demonstrating the feasibility of the PAM4 switched-capacitor lumped MZM for intra-data-centre links. Receiver-side digital equalisation is then employed to increase the reach of the link to 40 km, demonstrating extended potential for inter-data-centre links. As the experimental demonstration of this link was performed with optical amplification and a boosted extinction ratio of 7 dB (obtained by operating the MZM below its quadrature point), subsequent simulations are performed to analyse its performance in a non-optically amplified setting. Despite a reduction of approximately 2 dB in the extinction ratio, a moderately doped MZM held at

quadrature is shown to still operate successfully below FEC BER threshold for distances beyond 10 km.

As an alternative method of boosting lumped MZM bandwidths for data centre applications, electrical precompensation (transmitter-side equalisation, or predistortion) is also investigated. A simple one-tap FIR predistortion scheme is shown to enable 25 Gb/s NRZ modulation for an unterminated TW-MZM (i.e., operating essentially as a lumped MZM) with 5 GHz electro-optic bandwidth. The same one-tap predistortion also enables 50 Gb/s EDB modulation with the same MZM when additional receiver-side equalisation is employed, achieving the LR8 power budget over 10 km of fibre when driven with peak-to-peak voltages of 3 V in each arm. PAM4 is also demonstrated using a transmitter-side zero-forcing equaliser for predistortion, and, when coupled with receiver-side equalisation, achieves the LR8 power budget over 10 km. Comparison is also made between the unterminated TW-MZM and an MZM that is more truly designed to have lumped electrodes; some evidence of error flooring is shown for this lumped MZM, indicating that the unterminated TW-MZM presents an advantage due its better high-frequency design of electrodes. This suggests that the high-speed characteristics (such as RF attenuation) of lumped MZM electrodes are still important, and should be considered in device design flow.

Subsequent optimisation of the predistortion schemes to use matched filters at the transmitter side for the lumped MZM removes the need for any equalisation at the receiver, with demonstration of clear and open 50 Gb/s PAM4 and EDB eyes. Simulations are then used to evaluate the lumped MZM's performance over a range of fibre lengths. EDB outperforms PAM4 in terms of receiver sensitivity by 1.2 dB in a back-to-back link. However, PAM4 demonstrates more resilience to chromatic dispersion, achieving better sensitivity beyond 5 km. For a 2 V peak-to-peak drive signal in each arm, 50 Gb/s PAM4 comfortably reaches the KP4-FEC threshold over 10 km fibre, while EDB fails for distances beyond 8.5 km.

In order to evaluate the impact of predistortion on the drive voltages and energy consumption, an equivalent circuit model for the lumped MZM is developed. Fitting of the model's circuit parameters is performed using measured S11 parameters. The S21 frequency response of the circuit model is then compared against the measured frequency response of the lumped MZM, and sufficient agreement is found to justify the use of the model. Through this circuit model, comparison is made between terminated MZMs and unterminated MZMs, demonstrating clearly how the presence of a 50 Ω termination resistor extends

the bandwidth of the device but causes the modulation amplitude to drop by half. The unterminated lumped MZM is shown to be approximately 13 times more energy efficient than a terminated MZM at 10 Gbaud under drive conditions that achieve the same OMA.

Comparisons at higher baud rates are made with matched-filtering employed for the lumped MZM. For 25 Gbaud NRZ and PAM4 signal formats, the precompensation is shown to increase the required drive voltage by approximately a factor of 2 for the lumped MZM to achieve the same ER as before precompensation. This places its drive voltage requirements on par with a terminated MZM, although the lumped MZM's energy dissipation is still lower due to the absence of the termination resistor.

The circuit model is also used to compare the effects of using precompensation for lumped MZMs with the aforementioned alternative of using low-impedance drivers. The model shows that reducing the source impedance of the driver significantly improves the electrical bandwidth of the MZM from 6 GHz for a 50 Ω case to 38 GHz for a 5 Ω driver, with the bandwidth limited here by parasitic circuit elements. Additionally, the use of inductive peaking is shown to relax the requirement to have ultra-low source impedance, with a moderate amount of inductance (300 pH) shown to enable the same system performance at a higher driver impedance of 25 Ω . Notably, the performance of the lumped MZM is shown to be equivalent in the case of precompensation and when the optimal source impedance is used, showing the potential for a well-designed driver with low impedance and controlled amounts of inductive peaking to reduce the need for transmitter-side precompensation.

5.2 Future Work

The steep trend in interconnect speeds of recent years has caused rapid changes in data centre link architecture. Over the last decade, MMF and VCSELs have been largely superseded by SMF and silicon photonic modulators in order to facilitate 100 Gb/s links. Already, future 400 Gb/s links are standardised and being developed. Standards for 800 Gb/s and 1.6 Tb/s links are following. Achieving these milestones through simple increases in symbol rate is difficult due to device bandwidth limitations, and, hence, more spectrally efficient multilevel amplitude modulation formats such as PAM4 are likely to play a role. The switched-capacitor drive scheme studied in this thesis can be readily extended to

higher-order PAM formats (e.g., PAM8) by simply adding more capacitors into the capacitor bank. However, beyond PAM4, the sensitivity penalties brought about by the severely reduced eye openings, coupled with the increased complexity of the necessary receiver circuitry, may hinder the benefits of this route.

As intensity modulation and direct detection reach practical capacity limits and begin to plateau, coherent transmission may find a role in data centre links. By granting access to the phase of the signal as well the magnitude, coherent modulation schemes such as quadrature amplitude modulation (QAM) present an additional degree of freedom for encoding multiple bits per symbol. In addition, knowledge of the phase in the received signal allows receiver-side equalisation algorithms to significantly improve transmission distances over fibre. The availability of low-power and low-cost coherent transmission technology will be crucial for any adoption into data centre settings. Typically, parallel MZMs are used within an IQ modulator structure to generate coherent modulation formats. Switched-capacitor drivers could play a role here: banks of capacitor pairs driving each of the phase shifters in the four IQ modulator arms would facilitate 16-QAM. Higher order QAM could be reached by using additional capacitors in the banks.

Regardless of the chosen modulation format, modulator bandwidth and efficient use of drive voltage will remain a key priority. As symbol rates rise, silicon photonic TW-MZMs are likely to be preferred to lumped MZMs due to their superior bandwidths. Lumped MZM bandwidths (and lengths) are limited by the optical transit time of light through the device—an effect that can be removed through velocity-matching of the electrical and optical signals in TW-MZM design. Indeed, at very high frequencies, the presence of a termination resistor would be a relatively small price to pay in comparison to the power consumption of other driver overheads. However, alternative silicon photonic phase shifter technologies may prove to be key enablers for the use of lumped MZMs. Silicon-organic hybrid (SOH) integration combines silicon-on-insulator waveguides with electro-optic organic cladding materials in scalable post-processing step, and facilitates very short phase shifters with low V_π [145, 146]. Using such short-length phase shifters with switched-capacitor drivers or precompensation would be ideal for realising ultra high-speed lumped MZMs into the future.

Appendix A

BER of an EDB Signal

The following is a derivation of an approximate analytic expression for the BER of an EDB signal, as used in Equation 1.20. For legibility, in this section, the three duobinary symbols will be denoted 0, 1 and 2 rather than $-1, 0$ and 1 . As shown in Figure A.1, under thermal noise conditions, the photocurrent generated by the symbol samples at the receiver can be assumed to follow three normal distributions, with respective variances σ_0 , σ_1 and σ_2 , and respective means I_0 , I_1 and I_2 . The receiver uses two decision thresholds, denoted I_{thA} and I_{thB} .

Let $p(0)$, $p(1)$ and $p(2)$ be, respectively, the probabilities of symbol 0, 1 or 2 being transmitted. Let $P(1|0)$ be the probability of detecting a transmitted symbol 1 incorrectly as a symbol 0, with corresponding meanings given to $P(0|1)$, $P(1|2)$, $P(2|1)$, $P(0|2)$ and $P(2|0)$. Then the BER can be written as

$$BER = p(0)P(1|0) + p(1)[P(0|1) + P(2|1)] + p(2)P(1|2). \quad (A.1)$$

In this expression, $P(0|2)$ and $P(2|0)$ do not feature, as the bottom and top symbols of a duobinary sequence correspond to the same bit. Therefore, incorrectly detecting one as the other would not result in an error. Additionally,

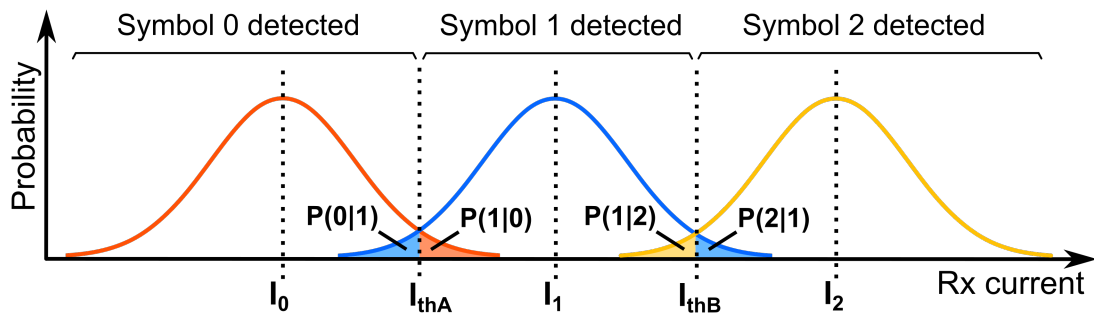


Figure A.1: Normal distribution of duobinary symbol samples and the receiver decision regions.

the mapping of a binary sequence to a duobinary sequence means the distribution of symbols gives $p(0) = p(2) = 1/4$ and $p(1) = 1/2$. This can be seen in the example duobinary mapping for PRBS4 given in Table 1.2.

As they follow normal distributions, the above conditional probabilities can be expressed as

$$P(0|1) = \frac{1}{\sigma_1\sqrt{2\pi}} \int_{-\infty}^{I_{thA}} \exp\left(-\frac{(I - I_1)^2}{2\sigma_1^2}\right) dI, \quad (\text{A.2})$$

$$\begin{aligned} P(1|0) &= \frac{1}{\sigma_0\sqrt{2\pi}} \int_{I_{thA}}^{I_{thB}} \exp\left(-\frac{(I - I_0)^2}{2\sigma_0^2}\right) dI \\ &= \frac{1}{\sigma_0\sqrt{2\pi}} \left[\int_{I_{thA}}^{\infty} \exp\left(-\frac{(I - I_0)^2}{2\sigma_0^2}\right) dI - \int_{I_{thB}}^{\infty} \exp\left(-\frac{(I - I_0)^2}{2\sigma_0^2}\right) dI \right], \end{aligned} \quad (\text{A.3})$$

$$\begin{aligned} P(1|2) &= \frac{1}{\sigma_2\sqrt{2\pi}} \int_{I_{thA}}^{I_{thB}} \exp\left(-\frac{(I - I_2)^2}{2\sigma_2^2}\right) dI \\ &= \frac{1}{\sigma_2\sqrt{2\pi}} \left[\int_{-\infty}^{I_{thB}} \exp\left(-\frac{(I - I_2)^2}{2\sigma_2^2}\right) dI - \int_{-\infty}^{I_{thA}} \exp\left(-\frac{(I - I_2)^2}{2\sigma_2^2}\right) dI \right], \end{aligned} \quad (\text{A.4})$$

$$P(2|1) = \frac{1}{\sigma_1\sqrt{2\pi}} \int_{I_{thA}}^{\infty} \exp\left(-\frac{(I - I_1)^2}{2\sigma_1^2}\right) dI. \quad (\text{A.5})$$

The change of integration limits in the above expressions for $P(1|0)$ and $P(1|2)$ was done to put the expressions in a suitable form to use the complimentary error function in the next step. This naturally introduces the subtracted terms.

Assuming that the noise variances are the same for each symbol level, i.e., $\sigma_0 = \sigma_1 = \sigma_2$, and introducing the complimentary error function as defined in Equation 1.10, these expressions for the conditional probabilities can be condensed to

$$P(0|1) = \frac{1}{2} \operatorname{erfc}\left(\frac{I_1 - I_{thA}}{\sigma\sqrt{2}}\right), \quad (\text{A.6})$$

$$P(1|0) = \frac{1}{2} \left[\operatorname{erfc}\left(\frac{I_{thA} - I_0}{\sigma\sqrt{2}}\right) - \operatorname{erfc}\left(\frac{I_{thB} - I_0}{\sigma\sqrt{2}}\right) \right], \quad (\text{A.7})$$

$$P(1|2) = \frac{1}{2} \left[\operatorname{erfc}\left(\frac{I_2 - I_{thB}}{\sigma\sqrt{2}}\right) - \operatorname{erfc}\left(\frac{I_2 - I_{thA}}{\sigma\sqrt{2}}\right) \right], \quad (\text{A.8})$$

$$P(2|1) = \frac{1}{2} \operatorname{erfc}\left(\frac{I_{thB} - I_1}{\sigma\sqrt{2}}\right). \quad (\text{A.9})$$

In addition, assuming that no symbols stray more than one symbol distance away from their correct position in terms of photocurrent, the subtracted terms in the expressions for $P(1|0)$ and $P(1|2)$ can be ignored, as shown. Plugging the

relevant expressions into Equation A.1, we arrive at

$$BER = \frac{1}{4} \left[\operatorname{erfc} \left(\frac{I_1 - I_{thA}}{\sigma \sqrt{2}} \right) + \operatorname{erfc} \left(\frac{I_{thB} - I_1}{\sigma \sqrt{2}} \right) \right] + \frac{1}{8} \left[\operatorname{erfc} \left(\frac{I_{thA} - I_0}{\sigma \sqrt{2}} \right) + \operatorname{erfc} \left(\frac{I_2 - I_{thB}}{\sigma \sqrt{2}} \right) \right]. \quad (\text{A.10})$$

As described for the NRZ derivation, the optimal threshold in a system limited by thermal noise is in the middle of any two levels. In this case, the resulting equal spacing of the thresholds means that the numerators of the erfc terms in Equation A.10 are equal, i.e., $I_{thA} - I_0 = I_1 - I_{thA} = I_{thB} - I_1 = I_2 - I_{thB}$. Introducing the average photoreceiver current I_{avg} , we then arrive at a convenient expression for the BER of an EDB signal:

$$BER = \frac{3}{4} \operatorname{erfc} \left(\frac{I_{avg}}{2\sqrt{2}\sigma} \right). \quad (\text{A.11})$$

Appendix B

Sinusoidal Current and Voltage in an RC Circuit

Lumped MZMs are often modelled after simple RC circuits, where the resistance R is given by the net combination of the driver's output resistance and the junction resistance of the phase shifter, and the capacitance C is the phase shifter's junction capacitance. The following derivation gives the frequency responses of the current and voltages present in such a circuit. These are used particularly in Section 4.3.1 for a discussion of the frequency dependence of lumped MZM energy consumption, along with initial examples of precompensation design for the lumped MZM.

Figure B.1 shows a diagram of an RC circuit, labelling the sinusoidal input voltage as $V_{in}(\omega)$, the current drawn as $I(\omega)$, and the voltage dropped across the capacitor as $V_{out}(\omega)$. The complex impedances of the resistor and capacitor can be represented respectively through $Z_R = R$ and $Z_{load} = 1/(i\omega C)$. The net

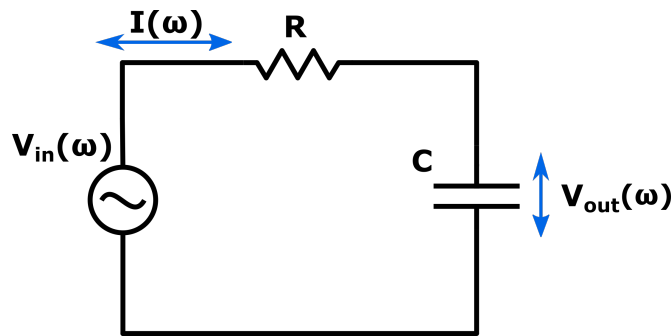


Figure B.1: A simple RC circuit diagram, which is often used to model the frequency behaviour of lumped MZMs.

impedance of the series RC combination is then given by $Z_{total} = Z_R + Z_{load}$. From here, the current drawn in the circuit is given by

$$I(\omega) = V_{in}(\omega) \cdot \frac{1}{Z_{total}} \quad (\text{B.1})$$

$$= V_{in}(\omega) \cdot \frac{1}{R - i/(\omega C)} \quad (\text{B.2})$$

$$= V_{in}(\omega) \cdot \frac{1}{|R - i/(\omega C)|} \cdot e^{i\phi} \quad (\text{B.3})$$

$$= V_{in}(\omega) \cdot \frac{1}{\sqrt{R^2 + 1/(\omega C)^2}} \cdot e^{i\phi}. \quad (\text{B.4})$$

The phase angle $\phi(\omega)$ comes out when changing the complex expression $I(\omega)$ to exponential form, and is given by

$$\phi(\omega) = \tan^{-1}(1/\omega RC). \quad (\text{B.5})$$

The frequency response of the current drawn by the RC circuit is thus given by

$$H_i(\omega) = \frac{1}{\sqrt{R^2 + 1/(\omega C)^2}} \cdot e^{i\phi}. \quad (\text{B.6})$$

From the voltage divider rule, the voltage dropping across the capacitor C is given by

$$V_{out}(\omega) = V_{in}(\omega) \cdot \frac{Z_{load}}{Z_{total}} \quad (\text{B.7})$$

$$= V_{in}(\omega) \cdot \frac{-i/(\omega C)}{R - i/(\omega C)} \quad (\text{B.8})$$

$$= V_{in}(\omega) \cdot \frac{1/(\omega C)}{|R - i/(\omega C)|} \cdot e^{-i\frac{\pi}{2}} \cdot e^{i\phi} \quad (\text{B.9})$$

$$= V_{in}(\omega) \cdot \frac{1}{\sqrt{1 + (\omega RC)^2}} \cdot e^{i\left(\phi(\omega) - \frac{\pi}{2}\right)} \quad (\text{B.10})$$

where the phase angle $\phi(\omega)$ is again given by Equation B.5. Hence the frequency response of the load voltage across the capacitor in an RC circuit is

$$H_v(\omega) = \frac{1}{\sqrt{1 + (\omega RC)^2}} \cdot e^{i\left(\phi(\omega) - \frac{\pi}{2}\right)}. \quad (\text{B.11})$$

Examining Equation B.6, it is clear that the current leads the input voltage by

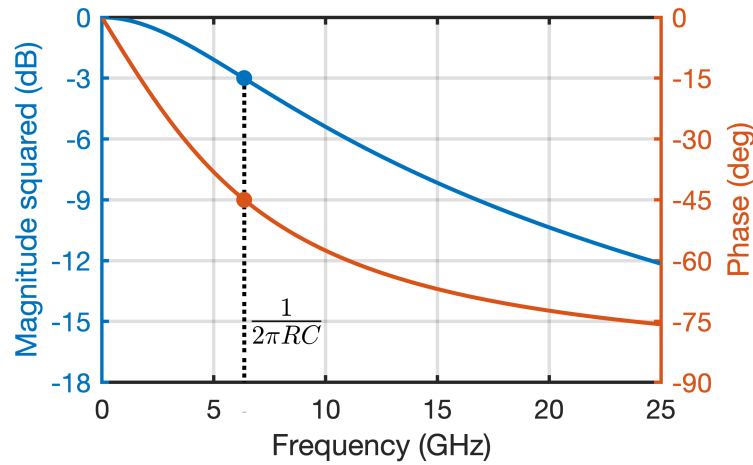


Figure B.2: RC frequency response ($R = 50 \, \Omega$, $C = 500 \, \text{fF}$).

a frequency-dependent angle $\phi(\omega)$ given by Equation B.5. Additionally, the presence of the $\exp(-i\pi/2)$ term in Equation B.11 gives the usual result that the voltage dropped across the capacitor in an RC circuit lags the current by $\pi/2$ radians.

Furthermore, from the equations, it can be seen that the capacitor acts as a short at high frequencies, as the current tends to V/R . Essentially, an RC circuit begins to look purely resistive under high-frequency modulation as less voltage is dropped across the capacitor and more is dropped across the resistor. The 3 dB bandwidth of the circuit is given by the frequency at which the power of the RF signal drops by a factor of 2, i.e., when magnitude of the RF voltage across the capacitor decreases by a factor of $\sqrt{2}$. This occurs at $\omega = 1/RC$.

Appendix C

Thévenin Equivalent Circuit Derivation for Lumped MZM

In Section 4.5.2, a Thévenin equivalent circuit representation of the lumped MZM was used to simplify notation. The following is a derivation of that equivalent circuit, showing how Equations 4.13 to 4.16 are obtained.

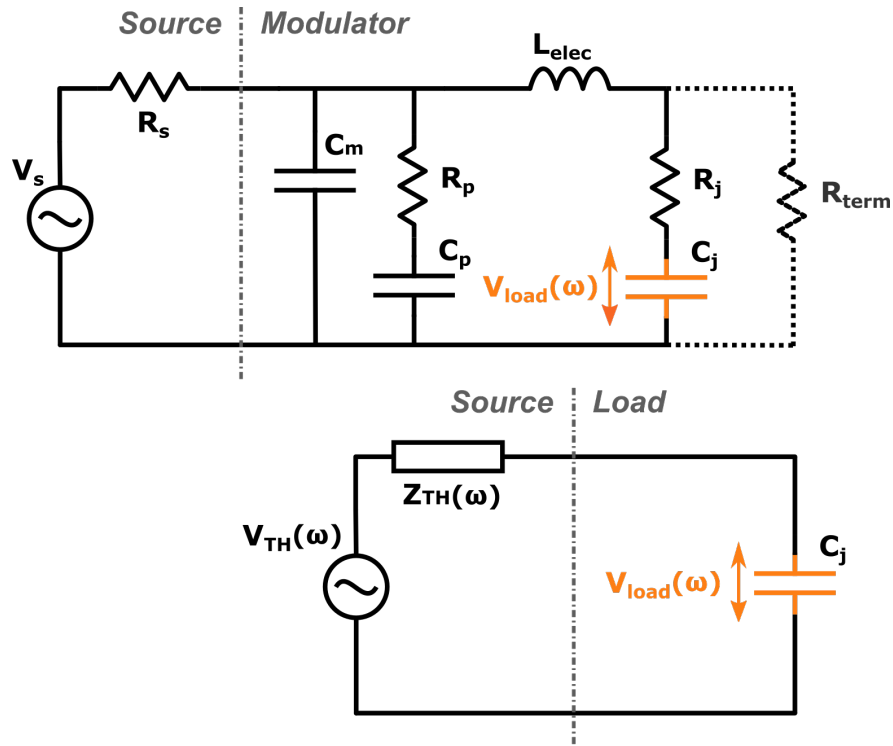


Figure C.1: The lumped MZM's equivalent circuit model and its Thévenin equivalent.

Thévenin's theorem, as applied to AC circuits, allows any linear circuit containing

only voltage sources, current sources and impedances to be replaced by an equivalent combination of a single voltage source V_{TH} in series with an impedance Z_{TH} . Essentially, it allow a circuit of many elements to be simplified to a circuit of just three elements, giving a source voltage and source impedance as seen from the “perspective” of a given element in the circuit. In this case, we desire the transfer function of the junction capacitance C_j , which gives the frequency response of the MZM, so the Thévenin circuit will be formed with respect to C_j , as shown in Figure C.1. The various elements of the lumped MZM’s equivalent circuit, along with their corresponding complex impedances, are described in detail in Section 4.5.1.

The first step of the Thévenin circuit derivation involves opening the load element C_j of the circuit at the point AB, as shown in Figure C.2, and calculating the resulting open circuit voltage. This is the Thévenin voltage, obtained after appropriate summing of impedances in series and parallel as

$$V_{TH}(\omega) = V_s(\omega) \cdot \frac{Z_1}{Z_2} \cdot \frac{Z_{Rt}}{Z_{Lelec} + Z_{Rt}}, \quad (C.1)$$

where we have set $Z_1 = Z_{Cm} \parallel (Z_{Rp} + Z_{Cp}) \parallel (Z_{Lelec} + Z_{Rt})$ and $Z_2 = Z_{Rs} + Z_1$ in order to simplify the notation. The junction resistance R_j does not appear in the calculation because the circuit is open at the point AB; no current flows through that branch. The $\parallel$ symbol indicates the usual calculation of impedances in parallel, i.e., two parallel impedances Z_x and Z_y give a combined impedance Z_{par} via $\frac{1}{Z_{par}} = \frac{1}{Z_x} + \frac{1}{Z_y}$.

The second step requires shorting the voltage source V_s (while keeping the open circuit at the section AB from the first step) and calculating the open circuit impedance at the point AB. This is the Thévenin impedance, and is obtained as

$$Z_{TH}(\omega) = \left([Z_{Rs} \parallel Z_{Cm} \parallel (Z_{Rp} + Z_{Cp}) + Z_{Lelec}] \parallel Z_{Rt} \right) + Z_{Rj}. \quad (C.2)$$

The junction resistance R_j does appear in this calculation as it presents an impedance in series with the rest of the combined elements of the circuit.

The final Thévenin circuit is made as shown in Figure C.2 by combining the Thévenin voltage V_{TH} and Thévenin impedance Z_{TH} in series with the junction capacitance C_j . Using the voltage divider rule, the desired load voltage across

C. THÉVENIN EQUIVALENT CIRCUIT
DERIVATION FOR LUMPED MZM

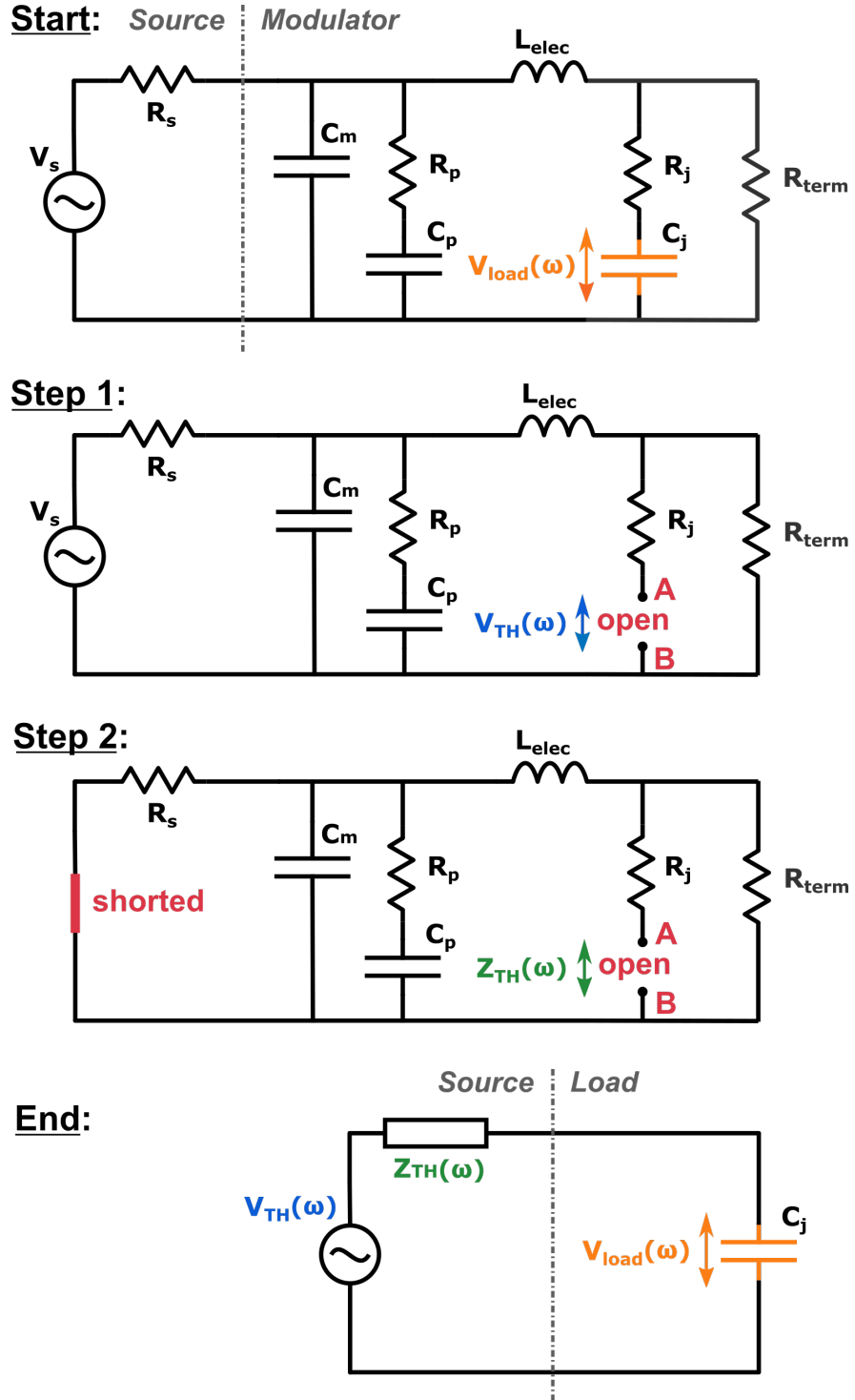


Figure C.2: Process for obtaining the Thévenin equivalent circuit used in Chapter 4 of this thesis.

the junction capacitance is then

$$V_{load}(\omega) = V_{TH}(\omega) \cdot \frac{Z_{Cj}}{Z_{TH}(\omega) + Z_{Cj}}. \quad (C.3)$$

The transfer function for the junction capacitor that determines the MZM frequency response is then obtained as

$$H_{vload}(\omega) = \frac{V_{load}(\omega)}{V_s(\omega)}. \quad (C.4)$$

Appendix D

A Brief Exhibition

Over the course of this PhD, much time was spent staring at modulators through a microscope camera while precisely aligning RF probes onto tiny metal pads. The following is a selection of photographs snapped during experiment setups, included here to give the interested reader a view beyond the diagram.

Although this thesis does not focus on VCSELs, the early part of my PhD work did involve some investigations into pushing VCSEL and MMF link bitrates using PAM4. Figure D.1 shows a picture of an 850 nm VCSEL taken during testing.

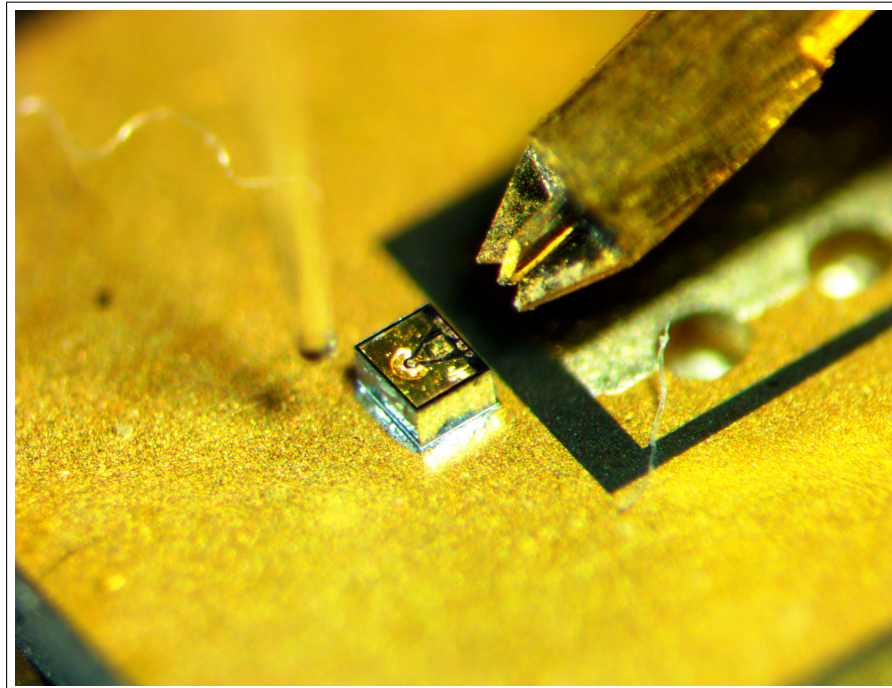


Figure D.1: Microscope picture of a VCSEL, vertically positioned optical fibre and RF probe.

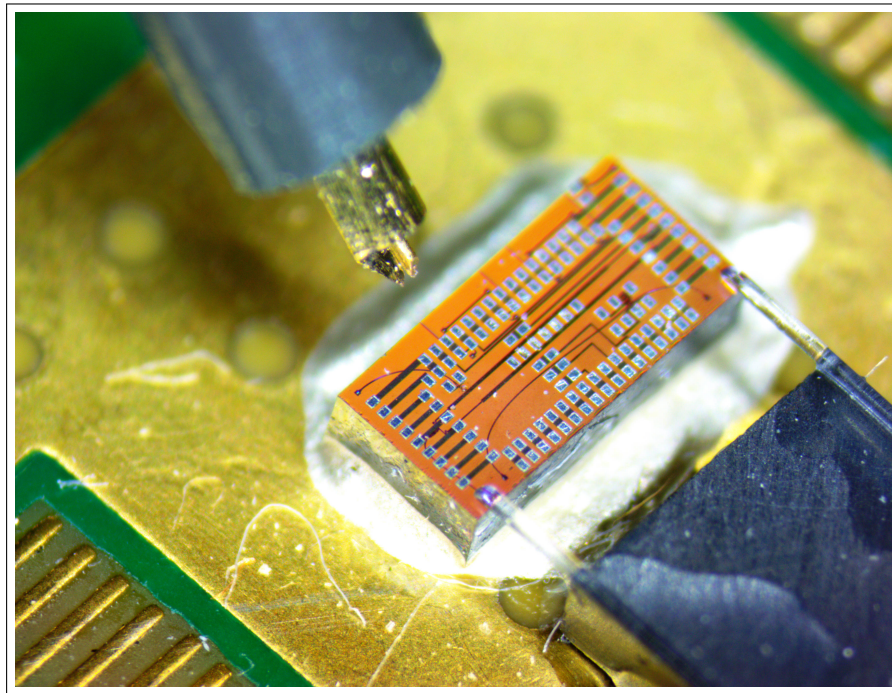


Figure D.2: Microscope picture of a lumped MZM used for measurements in this thesis, with RF probe and optical fibres visible.

A cleaved optical fibre hangs to the left, soon to be positioned over the central aperture of the VCSEL to capture output light. A high-speed RF probe looms from the right of frame, ready to make contact and drive current through the device. For a sense of scale, the three prongs (ground-signal-ground) are separated by just $125\text{ }\mu\text{m}$. Two small stray strands of dust feature in the image, one caught and twirled on the optical fibre, another lying to the right of the VCSEL on the metal substrate.

One of the lumped MZMs used for measurements in this thesis is shown in Figure D.2. An RF probe and planar-coupled optical fibres frame a silicon chip covered with myriad electrical contact pads and thin waveguides. The sharp-eyed reader will notice the thin dark line of the rib waveguide running from each optical fibre before forking into the interferometric structure of the MZM around the row of seven central RF contact pads. The chip itself appears in striking orange, a result of light rays diffracting or undergoing thin-film interference on the top surface before striking the camera. The colour seen is dependent on the angle of reflection between the camera and chip, an iridescence notable in the triptych of Figure D.3, where the slight repositioning of the lenses when zooming causes a transition from deep blue to violet.

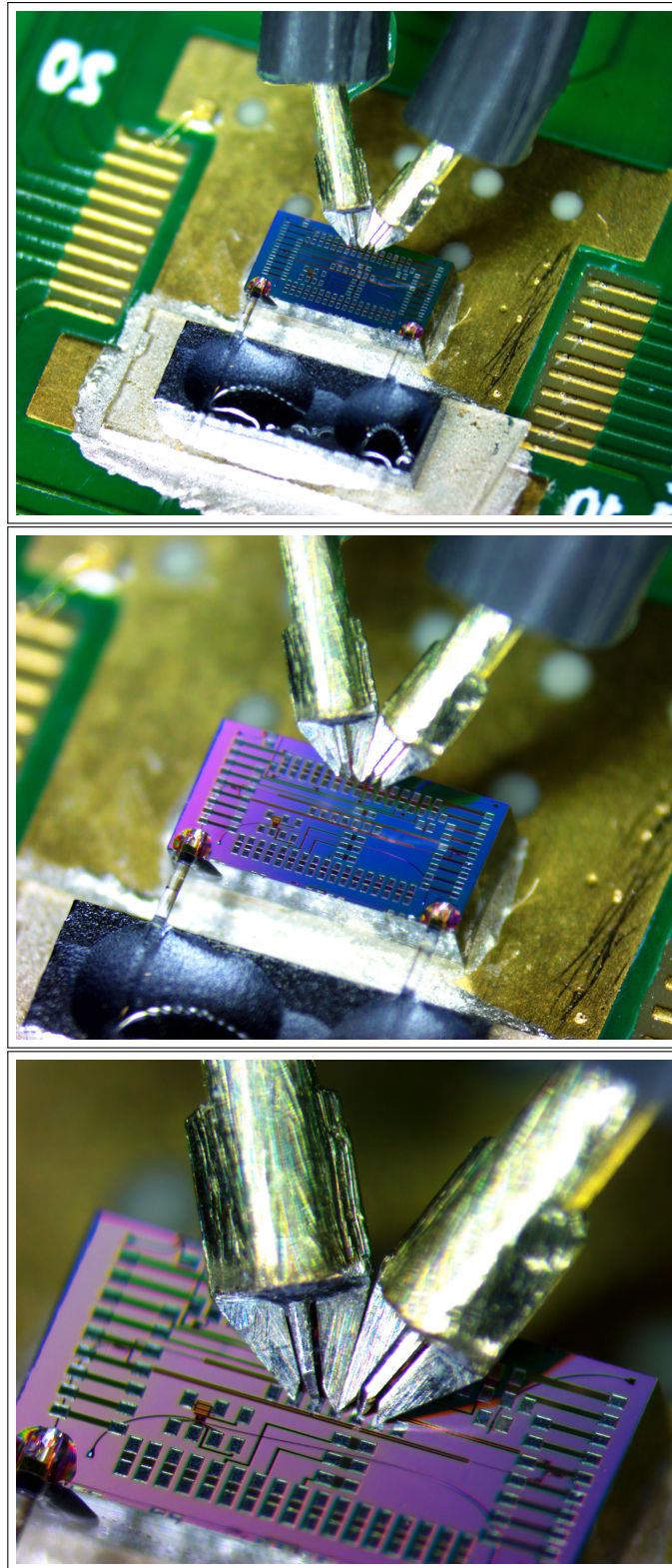


Figure D.3: Aligning RF probes with the lumped MZM, seen from three different levels of analogue zoom.

List of Acronyms and Initialisms

ASE	amplified spontaneous emission
AWG	arbitrary waveform generator
B2B	back to back
BER	bit error rate
BPF	bandpass filter
CAGR	compound annual growth rate
CMOS	complimentary metal-oxide semiconductor
CTLE	continuous-time linear equaliser
DAC	digital-analog converter
DSP	digital signal processing
EDB	electrical duobinary
EDC	electronic dispersion compensation
EDFA	erbium-doped fibre amplifier
ER	extinction ratio
FEC	forward error correction
FEM	finite element method
FFE	feed-forward equaliser
FIR	finite impulse response
FSR	free spectral range
FTTB	fibre to the building

FTTH	fibre to the home
HTTP	hypertext transfer protocol
I/O	input-output
IL	insertion loss
ISI	intersymbol interference
LAN	local area network
LMR	level mismatch ratio
LSB	least significant bit
MSB	most significant bit
MSE	mean squared error
MZI	Mach-Zehnder interferometer
MZM	Mach-Zehnder modulator
NRZ	non return to zero
OMA	optical modulation amplitude
OOK	on-off keying
PAM	pulse amplitude modulation
PAM2	two-level pulse-amplitude modulation
PAM4	four-level pulse amplitude modulation
PAM8	eight-level pulse amplitude modulation
PC	polarisation controller
PIC	photonic integrated circuit
PPG	pulse pattern generator
PRBS	pseudo-random binary sequence
QAM	quadrature amplitude modulation
RF	radio frequency
RTO	real-time oscilloscope
RTO	real-time oscilloscope

LIST OF ACRONYMS AND INITIALISMS

Rx	receiver
SER	symbol error rate
SiP	silicon photonics
SMF	multimode fibre
SMF	single-mode fibre
SOH	silicon-organic hybrid
TCAD	technology computer-aided design
TIA	transimpedance amplifier
TLP	transmission link penalty
TOR	top of rack
TW-MZM	travelling-wave Mach-Zehnder modulator
Tx	transmitter
VCSEL	vertical-cavity surface-emitting laser
VNA	vector network analyser
VOA	variable optical attenuator
WAN	wide area network
WDM	wavelength-division multiplexing
XOR	exclusive-or

List of Publications

The following is a list of journal and conference contributions arising as a result of the research carried out during this PhD.

Published Journals and Conference Contributions

1. **B. Murray**, C. Antony, G. Talli, and P. D. Townsend, “Predistortion for High-Speed Lumped Silicon Photonic Mach-Zehnder Modulators”, in *IEEE Photonics Journal*, vol. 14, no. 2, pp. 1–11, April 2022.
2. **B. Murray**, C. Antony, G. Talli, and P. D. Townsend, “50-Gbps PAM4 and electrical duobinary modulation with a lumped silicon photonic MZM,” in *45th European Conference on Optical Communication (ECOC 2019)*, pp. 1–4 2019.
3. M. Power, **B. Murray**, S. Facchin, S. Zhou, C. Antony, C. Eason, C. Scarcella, G. Talli, P. O’Brien, P. Ossieur, and P. D. Townsend, “Low power 15 Gbaud silicon photonic link for data centres”, in *2018 European Conference on Optical Communications (ECOC)*, pp. 1–3, 2017.

Unpublished Conference Contributions

1. **B. Murray**, C. Antony, G. Talli, P. D. Townsend, “Optimisation of High-Speed Lumped Silicon Photonic Modulators” in *Photonics Ireland*, Dublin, 2021.
2. **B. Murray**, C. Antony, G. Talli, P. D. Townsend, “Optimisation of Lumped PAM4/PAM8 Silicon Photonic Mach-Zehnder Modulators ”, in *Photonics Ireland*, Cork, 2018.
3. **B. Murray**, C. Antony, S. Duggan, G. Talli, P. D. Townsend, “Bias and Doping

Optimisation of Lumped PAM4/PAM8 Mach-Zehnder Modulators”, in *20th European Conference of Integrated Optics (ECIO)*, Valencia, 2018.

4. **B. Murray**, C. Antony, G. Talli, P. D. Townsend, “15 GBaud Silicon Photonic Interconnect.”, in *Photonics Ireland*, Galway, 2017.
5. **B. Murray**, C. Antony, G. Talli, P. D. Townsend, “PAM4 for Next-Generation VCSEL-Based Short-Reach Interconnects”, in *Photonics Ireland*, Cork, 2015.

References

- [1] Telegeography, “Submarine Cable Map,” 2020. <https://www.submarinecablemap.com>. Accessed 13 July 2020.
- [2] Cisco, “Annual Internet Report 2018-2023,” tech. rep., 2020.
- [3] R. Montagne, “FTTH Council Europe Conference: Europe Broadband Status,” 2019.
- [4] A. Holst, “Number of data centers in Europe by country 2019,” 2020. Available at <https://www.statista.com/statistics/878621/european-data-centers-by-country>. Accessed 28 June 2020.
- [5] Host in Ireland, “Ireland’s data hosting industry: 2019 Q4 report,” tech. rep., 2020.
- [6] Host in Ireland, “Ireland’s data hosting industry: 2020 Q1 report,” tech. rep., 2020.
- [7] N. Parfitt, “The Future Data Centre: trends in smart data centre design and construction,” tech. rep., Data Centre Dynamics Group and Enterprise Ireland, 2017.
- [8] D. Inniss and R. Rubenstein, *Silicon Photonics: Fueling the Next Information Revolution*. Cambridge, MA: Morgan Kaufmann, 2017.
- [9] J. G. Koomey, “Worldwide electricity used in data centers,” *Environmental research letters*, vol. 3, no. 3, pp. 220–236, 2008.
- [10] P. Corcoran and A. Andrae, “Emerging trends in electricity consumption for consumer ICT,” *National University of Ireland, Galway, Connacht, Ireland, Tech. Rep*, 2013.

- [11] M. Dayarathna, Y. Wen, and R. Fan, “Data Center Energy Consumption Modeling: A Survey,” *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 732–794, 2015.
- [12] A. S. Andrae and T. Edler, “On global electricity usage of communication technology: trends to 2030,” *Challenges*, vol. 6, no. 1, pp. 117–157, 2015.
- [13] L. Belkhir and A. Elmeligi, “Assessing ICT global emissions footprint: Trends to 2040 & recommendations,” *Journal of Cleaner Production*, vol. 177, pp. 448 – 463, 2018.
- [14] E. Masanet, A. Shehabi, N. Lei, S. Smith, and J. Koomey, “Recalibrating global data center energy-use estimates,” *Science*, vol. 367, no. 6481, pp. 984–986, 2020.
- [15] D. A. B. Miller, “Device Requirements for Optical Interconnects to Silicon Chips,” *Proceedings of the IEEE*, vol. 97, no. 7, pp. 1166–1185, 2009.
- [16] N. Farrington and A. Andreyev, “Facebook’s data center network architecture,” in *2013 Optical Interconnects Conference*, pp. 49–50, 2013.
- [17] A. Andreyev, X. Wang, and A. Eckert, “Reinventing Facebook’s data center network,” Mar. 2019. Available at <https://engineering.fb.com/data-center-engineering/f16-minipack>. Accessed on 15 July 2020.
- [18] C. Spurgeon, *Ethernet: The Definitive Guide*. O’Reilly Media, 2000.
- [19] R. Metcalfe and D. Boggs, “35 mm slide showing Ethernet diagram, presented to the National Computer Conference 1976.,” 1976. Available at http://www.ieee802.org/3/ethernet_diag.html. Accessed 1 July 2020.
- [20] R. M. Metcalfe and D. R. Boggs, “Ethernet: Distributed packet switching for local computer networks,” *Communications of the ACM*, vol. 19, no. 7, pp. 395–404, 1976.
- [21] R. Colburn, “Robert Metcalfe, an oral history conducted in 2004 by Robert Colburn.” IEEE History Center, Hoboken, NJ, USA, 2004. Available at https://ethw.org/Oral-History:Robert_Metcalfe. Accessed 25 November 2020.
- [22] B. Forouzan, *Data Communications and Networking*. McGraw-Hill Education, 2012.

- [23] Engineering and Technology History Wiki, “Transatlantic Optical Cable.” Webpage, 2017. Available at https://ethw.org/Transatlantic_Optical_Cable. Accessed 3 July 2020.
- [24] M. Sibley, *Optical Communications: Components and Systems*. Springer Nature, third ed., 2020.
- [25] Cisco, “Fiber Types in Gigabit Optical Communications,” tech. rep., Cisco, 2008.
- [26] Thor Labs, “Graded-Index (GRIN) Multimode Fibers, GIF50x Typical Spectral Absorption.” Data sheet, 2017. https://www.thorlabs.com/newgrouppage9.cfm?objectgroup_id=358. Accessed 25th June, 2020.
- [27] A. Rohatgi, “Webplotdigitizer: Version 4.3,” 2020. Available at <https://automeris.io/WebPlotDigitizer>.
- [28] Xerox Corporation, “The Ethernet: A Local Area Network: Data Link Layer and Physical Layer Specifications,” *SIGCOMM Comput. Commun. Rev.*, vol. 11, p. 20–66, July 1981.
- [29] C. DeCusatis, *Handbook of Fiber Optic Data Communication: A Practical Guide to Optical Networking*. Elsevier Science, 2013.
- [30] IEEE Ethernet Working Group, “802.3.” Available at <http://www.ieee802.org/3/>. Accessed 26 June 2020.
- [31] H. Lu, C. Li, C. Lu, C. Ho, H. Chen, M. Cheng, Z. Yang, and S. Huang, “A 103.12-Gb/s WDM PAM4 VCSEL-Based Transmission With Light Injection and Optoelectronic Feedback Techniques,” *IEEE Photonics Journal*, vol. 9, no. 1, pp. 1–8, 2017.
- [32] K. Panajotov and R. Schatz, “Coupled-Cavity VCSEL with an Integrated Electro-Absorption Modulator: Small and Large-Signal Modulation Analysis,” *Applied Sciences*, vol. 10, no. 17, p. 6128, 2020.
- [33] R. Soref and B. Bennett, “Kramers-Kronig Analysis of Electro-Optical Switching in Silicon,” *Proc. SPIE*, vol. 704, 01 1986.
- [34] R. Soref and B. Bennett, “Electrooptical effects in silicon,” *IEEE Journal of Quantum Electronics*, vol. 23, pp. 123–129, January 1987.
- [35] R. A. Soref, “Silicon-based optoelectronics,” *Proceedings of the IEEE*, vol. 81, no. 12, pp. 1687–1706, 1993.

- [36] B. Szelag, K. Hassan, L. Adelmini, K. Ribaud, M. C. Roure, L. Sanchez, A. Schembri, and C. Jany, “CMOS Compatible III-V/Silicon Laser Integration on Silicon Photonics Platform,” in *2019 IEEE Photonics Conference (IPC)*, pp. 1–2, IEEE, 2019.
- [37] N. Li, M. Xin, Z. Su, E. S. Magden, N. Singh, J. Notaros, E. Timurdogan, P. Purnawirman, J. D. B. Bradley, and M. R. Watts, “A silicon photonic data link with a monolithic erbium-doped laser,” *Scientific reports*, vol. 10, no. 1, pp. 1–9, 2020.
- [38] A. C. Clarke, *How the World Was One: Beyond the Global Village*. Gollancz, 1992.
- [39] G. P. Agrawal, *Fiber-Optic Communication Systems*. John Wiley and Sons, fourth ed., 2010.
- [40] K. Szczerba, P. Westbergh, J. Karout, J. S. Gustavsson, A. Haglund, M. Karlsson, P. A. Andrekson, E. Agrell, and A. Larsson, “4-PAM for high-speed short-range optical communications,” *IEEE/OSA Journal of Optical Communications and Networking*, vol. 4, no. 11, pp. 885–894, 2012.
- [41] P. Madsen, *Partial Response Advanced Modulation Formats for Bandwidth Limited Optical Links*. PhD thesis, Technical University of Denmark, 2019.
- [42] G. T. Reed and A. P. Knights, *Silicon photonics: an introduction*. John Wiley & Sons, 2004.
- [43] L. Chrostowski and M. Hochberg, *Silicon Photonics Design*. Cambridge University Press, 2015.
- [44] D. Patel, *Design, analysis, and performance of a silicon photonic traveling wave Mach-Zehnder modulator*. McGill University (Canada), 2015.
- [45] A. A. Michelson and E. W. Morley, “LVIII. On the relative motion of the earth and the luminiferous æther,” *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 24, no. 151, pp. 449–463, 1887.
- [46] R. S. Shankland, “Michelson-Morley Experiment,” *American Journal of Physics*, vol. 32, no. 1, pp. 16–35, 1964.
- [47] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, *et al.*,

- “Observation of gravitational waves from a binary black hole merger,” *Physical review letters*, vol. 116, no. 6, p. 061102, 2016.
- [48] Synopsys, “Femsim.” Software, 2020. <https://www.synopsys.com/photonic-solutions/rsoft-photonic-device-tools/passive-device-femsim.html>. Accessed 8 May 2020.
- [49] H. Huang, K. Liu, B. Qi, and V. J. Sorger, “Re-analysis of single-mode conditions for silicon rib waveguides at 1550 nm wavelength,” *Journal of Lightwave Technology*, vol. 34, no. 16, pp. 3811–3817, 2016.
- [50] F. Payne and J. Lacey, “A theoretical analysis of scattering loss from planar optical waveguides,” *Optical and Quantum Electronics*, vol. 26, no. 10, pp. 977–986, 1994.
- [51] K. K. Lee, D. R. Lim, L. C. Kimerling, J. Shin, and F. Cerrina, “Fabrication of ultralow-loss si/sio₂ waveguides by roughness reduction,” *Opt. Lett.*, vol. 26, pp. 1888–1890, Dec 2001.
- [52] E. Dulkeith, F. Xia, L. Schares, W. M. Green, and Y. A. Vlasov, “Group index and group velocity dispersion in silicon-on-insulator photonic wires,” *Optics express*, vol. 14, no. 9, pp. 3853–3863, 2006.
- [53] S. Dwivedi, T. Van Vaerenbergh, A. Ruocco, T. Spuesens, P. Bienstman, P. Dumon, and W. Bogaerts, “Measurements of effective refractive index of soi waveguides using interferometers,” in *Integrated Photonics Research, Silicon and Nanophotonics*, pp. IM2A–6, Optical Society of America, 2015.
- [54] L. Cheng, S. Mao, Z. Li, Y. Han, and H. Fu, “Grating couplers on silicon photonics: Design principles, emerging trends and practical issues,” *Micromachines*, vol. 11, no. 7, p. 666, 2020.
- [55] B. Snyder and P. O’Brien, “Planar fiber packaging method for silicon photonic integrated circuits,” in *OFC/NFOEC*, pp. 1–3, 2012.
- [56] L. Zheng, S. Shao, J. Ding, L. Zhang, X. Fu, and L. Yang, “Intrinsic chirp analysis of a single dual-drive silicon pam-4 optical modulator,” *Optics Express*, vol. 26, pp. 32014–32025, Nov 2018.
- [57] P. Dong, L. Chen, and Y.-k. Chen, “High-speed low-voltage single-drive push-pull silicon mach-zehnder modulators,” *Optics express*, vol. 20, no. 6, pp. 6163–6169, 2012.

- [58] D. Patel, S. Ghosh, M. Chagnon, A. Samani, V. Veerasubramanian, M. Osman, and D. V. Plant, "Design, analysis, and transmission system performance of a 41 ghz silicon photonic modulator," *Optics Express*, vol. 23, pp. 14263–14287, Jun 2015.
- [59] M. Ker, C. Lin, and Y. Hsiao, "Overview on esd protection designs of low-parasitic capacitance for rf ics in cmos technologies," *IEEE Transactions on Device and Materials Reliability*, vol. 11, pp. 207–218, June 2011.
- [60] G. Boselli, J. Rodriguez, C. Duvvury, and J. Smith, "Analysis of esd protection components in 65nm cmos technology: Scaling perspective and impact on esd design window," in *2005 Electrical Overstress/Electrostatic Discharge Symposium*, pp. 1–10, Sep. 2005.
- [61] G. P. Agrawal and N. K. Dutta, *Semiconductor Lasers*. New York: Van Nostrand Reinhold, second ed., 1993.
- [62] S. K. Selvaraja, W. Bogaerts, P. Dumon, D. Van Thourhout, and R. Baets, "Subnanometer linewidth uniformity in silicon nanophotonic waveguide devices using cmos fabrication technology," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 16, pp. 316–324, Jan 2010.
- [63] G. Cocorullo and I. Rendina, "Thermo-optical modulation at 1.5 μm in silicon etalon," *Electronics Letters*, vol. 28, pp. 83–85, Jan 1992.
- [64] J. Witzens, "High-speed silicon photonics modulators," *Proceedings of the IEEE*, vol. 106, no. 12, pp. 2158–2182, 2018.
- [65] Multi-Source Agreement (MSA), "Quad small form-factor pluggable (qfsp) transceiver specification." SFF Committee, Dec. 2006.
- [66] R. Amatya, C. W. Holzwarth, M. A. Popovic, F. Gan, H. I. Smith, F. Kartner, and R. J. Ram, "Low power thermal tuning of second-order microring resonators," in *2007 Conference on Lasers and Electro-Optics (CLEO)*, pp. 1–2, 2007.
- [67] M. R. Watts, W. A. Zortman, D. C. Trotter, G. N. Nielson, D. L. Luck, and R. W. Young, "Adiabatic resonant microrings (arms) with directly integrated thermal microphotronics," in *2009 Conference on Lasers and Electro-Optics and 2009 Conference on Quantum electronics and Laser Science Conference*, pp. 1–2, IEEE, 2009.

- [68] B. Guha, A. Gondarenko, and M. Lipson, “Minimizing temperature sensitivity of silicon mach-zehnder interferometers,” *Opt. Express*, vol. 18, pp. 1879–1887, Feb 2010.
- [69] M. Leidinger, S. Fieberg, N. Waasem, F. Kühnemann, K. Buse, and I. Breunig, “Comparative study on three highly sensitive absorption measurement techniques characterizing lithium niobate over its entire transparent spectral range,” *Optics express*, vol. 23, no. 17, pp. 21690–21705, 2015.
- [70] R. Tilley, *Crystals and Crystal Structures*. Wiley, 2006.
- [71] C. Schriever, C. Bohley, J. Schilling, and R. B. Wehrspohn, “Strained silicon photonics,” *Materials*, vol. 5, no. 5, pp. 889–908, 2012.
- [72] T. S. Moss, G. Burrell, and B. Ellis, *Semiconductor Opto-Electronics*. London: Butterworths, 1973.
- [73] S. M. Sze and K. K. Ng, *Physics of semiconductor devices*. John Wiley & Sons, 2006.
- [74] F. Wooten, *Optical Properties of Solids*. North-Holland Publishing Company, 1972.
- [75] A. Chen and E. Murphy, *Broadband Optical Modulators: Science, Technology, and Applications*. Taylor & Francis, 2011.
- [76] S. Chung, M. Nakai, and H. Hashemi, “Low-power thermo-optic silicon modulator for large-scale photonic integrated systems,” *Opt. Express*, vol. 27, pp. 13430–13459, Apr 2019.
- [77] M. Nedeljkovic, R. Soref, and G. Z. Mashanovich, “Free-carrier electrorefraction and electroabsorption modulation predictions for silicon over the 1–14- μ m infrared wavelength range,” *IEEE Photonics Journal*, vol. 3, no. 6, pp. 1171–1180, 2011.
- [78] C. E. Png, M. J. Sun, S. T. Lim, T. Y. L. Ang, and K. Ogawa, “Numerical modeling and analysis for high-efficiency carrier-depletion silicon rib-waveguide phase shifters,” *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 22, pp. 99–106, Nov 2016.
- [79] G. T. Reed, G. Mashanovich, F. Y. Gardes, and D. J. Thomson, “Silicon optical modulators,” *Nature Photonics*, vol. 4, no. 8, pp. 518–526, 2010.

- [80] G. Ghione, *Semiconductor Devices for High-Speed Optoelectronics*. Cambridge University Press, 2009.
- [81] H. Johnson, H. W. Johnson, and M. Graham, *High-speed Signal Propagation: Advanced Black Magic*. Prentice Hall PTR Signal Integrity Library, Prentice Hall PTR, 2003.
- [82] T. L. Koch, F. J. Leonberger, and P. G. Suchoski, “Integrated optics,” in *Handbook of Optics, Vol. 1 - Components and Instruments* (M. Bass, ed.), vol. 1, ch. 6, McGraw-Hill, second, illustrated ed., 1995.
- [83] S. S. Azadeh, F. Merget, S. Romero-García, A. Moscoso-Mártir, N. von den Driesch, J. Müller, S. Mantl, D. Buca, and J. Witzens, “Low v_{π} silicon photonics modulators with highly linear epitaxially grown phase shifters,” *Optics express*, vol. 23, no. 18, pp. 23526–23550, 2015.
- [84] S. Zhou, H.-t. Wu, K. Sadeghipour, C. Scarcella, C. Eason, M. Rensing, M. J. Power, C. Antony, P. O’Brien, P. D. Townsend, *et al.*, “Optimization of pam-4 transmitters based on lumped silicon photonic mzms for high-speed short-reach optical links,” *Optics express*, vol. 25, no. 4, pp. 4312–4325, 2017.
- [85] H. Xu, X. Xiao, X. Li, Y. Hu, Z. Li, T. Chu, Y. Yu, and J. Yu, “High speed silicon mach-zehnder modulator based on interleaved pn junctions,” *Optics express*, vol. 20, no. 14, pp. 15093–15099, 2012.
- [86] X. Xiao, H. Xu, X. Li, Z. Li, T. Chu, Y. Yu, and J. Yu, “High-speed, low-loss silicon mach-zehnder modulators with doping optimization,” *Opt. Express*, vol. 21, pp. 4116–4125, Feb 2013.
- [87] M. Dummer, M. Sysak, A. Tauke-Pedretti, J. Raring, J. Klamkin, and L. Coldren, “Widely tunable separate absorption and modulation wavelength converter with integrated microwave termination,” *Lightwave Technology, Journal of*, vol. 26, pp. 938–944, 05 2008.
- [88] W. D. Jemison and A. C. Paolella, “Analog fiber-optic links,” in *The RF and Microwave Handbook - Volume 3: RF and Microwave Applications and Systems* (M. Golio, ed.), Electrical Engineering Handbook, ch. 9, CRC Press, second, revised ed., 2018.
- [89] M. Lauermann, *Silicon-organic hybrid devices for high-speed electro-optic signal processing*. Karlsruhe Series in Photonics and Communications /

- Karlsruhe Institute of Technology, Institute of Photonics and Quantum Electronics (IPQ), KIT Scientific Publishing, 2018.
- [90] L. Friedman, R. A. Soref, and J. P. Lorenzo, “Silicon double-injection electro-optic modulator with junction gate control,” *Journal of applied physics*, vol. 63, no. 6, pp. 1831–1839, 1988.
 - [91] A. Liu, R. Jones, L. Liao, D. Samara-Rubio, D. Rubin, O. Cohen, R. Nicolaescu, and M. Paniccia, “A high-speed silicon optical modulator based on a metal–oxide–semiconductor capacitor,” *Nature*, vol. 427, no. 6975, pp. 615–618, 2004.
 - [92] L. Liao, D. Samara-Rubio, M. Morse, A. Liu, D. Hodge, D. Rubin, U. D. Keil, and T. Franck, “High speed silicon mach-zehnder modulator,” *Opt. Express*, vol. 13, pp. 3129–3135, Apr 2005.
 - [93] F. Y. Gardes, G. T. Reed, N. Emerson, and C. Png, “A sub-micron depletion-type photonic modulator in silicon on insulator,” *Opt. Express*, vol. 13, pp. 8845–8854, Oct 2005.
 - [94] A. Liu, L. Liao, D. Rubin, H. Nguyen, B. Ciftcioglu, Y. Chetrit, N. Izhaky, and M. Paniccia, “High-speed optical modulation based on carrier depletion in a silicon waveguide,” *Opt. Express*, vol. 15, pp. 660–668, Jan 2007.
 - [95] L. Liao, A. Liu, D. Rubin, J. Basak, Y. Chetrit, H. Nguyen, R. Cohen, N. Izhaky, and M. Paniccia, “40 gbit/s silicon optical modulator for highspeed applications,” *Electronics Letters*, vol. 43, no. 22, 2007.
 - [96] F. Y. Gardes, D. J. Thomson, N. G. Emerson, and G. T. Reed, “40 gb/s silicon photonics modulator for te and tm polarisations,” *Opt. Express*, vol. 19, pp. 11804–11814, Jun 2011.
 - [97] D. Thomson, F. Gardes, Y. Hu, G. Mashanovich, M. Fournier, P. Grosse, J. Fedeli, and G. Reed, “High contrast 40gbit/s optical modulation in silicon,” *Optics express*, vol. 19, no. 12, pp. 11507–11516, 2011.
 - [98] D. J. Thomson, F. Y. Gardes, J. Fedeli, S. Zlatanovic, Y. Hu, B. P. P. Kuo, E. Myslivets, N. Alic, S. Radic, G. Z. Mashanovich, and G. T. Reed, “50-gb/s silicon optical modulator,” *IEEE Photonics Technology Letters*, vol. 24, no. 4, pp. 234–236, 2012.
 - [99] M. Ziebell, D. Marris-Morini, G. Rasigade, J.-M. Fédéli, P. Crozat, E. Cassan, D. Bouville, and L. Vivien, “40 gbit/s low-loss silicon optical

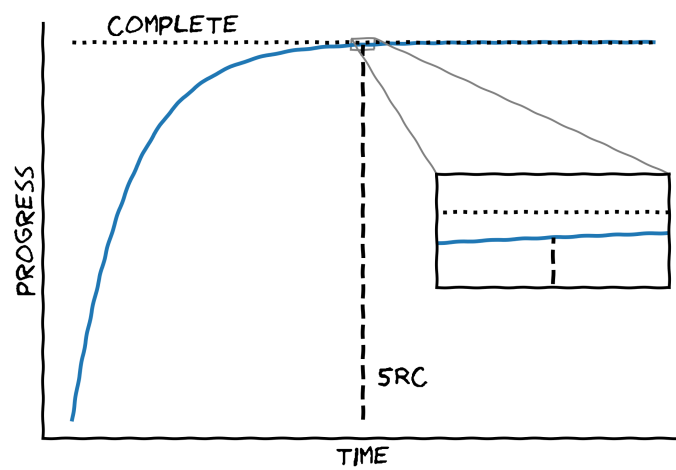
- modulator based on a pipin diode,” *Opt. Express*, vol. 20, pp. 10591–10596, May 2012.
- [100] M. Li, L. Wang, X. Li, X. Xiao, and S. Yu, “Silicon intensity mach-zehnder modulator for single lane 100 gb/s applications,” *Photon. Res.*, vol. 6, pp. 109–116, Feb 2018.
- [101] H. Yu, M. Pantouvaki, J. Van Campenhout, D. Korn, K. Komorowska, P. Dumon, Y. Li, P. Verheyen, P. Absil, L. Alloatti, D. Hillerkuss, J. Leuthold, R. Baets, and W. Bogaerts, “Performance tradeoff between lateral and interdigitated doping patterns for high speed carrier-depletion based silicon modulators,” *Optics Express*, vol. 20, pp. 12926–38, 06 2012.
- [102] A. Michard, J. Carpentier, N. Michit, P. Le Maître, P. Bénabès, and P. M. Ferreira, “A sub-pj/bit, low-er mach–zehnder-based transmitter for chip-to-chip optical interconnects,” *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 26, no. 2, pp. 1–10, 2020.
- [103] S. S. Azadeh, J. Nojić, A. Moscoso-Mártir, F. Merget, and J. Witzens, “Power-efficient lumped-element meandered silicon Mach-Zehnder modulators,” in *Silicon Photonics XV* (G. T. Reed and A. P. Knights, eds.), vol. 11285, pp. 65 – 75, International Society for Optics and Photonics, SPIE, 2020.
- [104] G. Cong, M. Ohno, Y. Maegami, M. Okano, R. Kou, N. Yamamoto, and K. Yamada, “Ultra-compact non-travelling-wave silicon mach-zehnder modulator,” in *45th European Conference on Optical Communication (ECOC 2019)*, pp. 1–3, 2019.
- [105] G. Cong, Y. Maegami, M. Ohno, and K. Yamada, “Ultra-compact non-travelling-wave silicon carrier-depletion mach-zehnder modulators towards high channel density integration,” *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 27, no. 3, pp. 1–11, 2021.
- [106] Y. Sobu, S. Tanaka, and Y. Tanaka, “High-speed-operation of all-silicon lumped-electrode modulator integrated with passive equalizer,” *IEICE Transactions on Electronics*, 05 2020.
- [107] S. O. Kasap, *Principles of electronic materials and devices*. McGraw-Hill, fourth ed., 2018.
- [108] K. Goi, K. Ogawa, Y. T. Tan, V. Dixit, S. T. Lim, C. E. Png, T.-Y. Liow, X. Tu, G.-Q. Lo, and D.-L. Kwong, “Silicon mach-zehnder modulator using low-

- loss phase shifter with bottom pn junction formed by restricted-depth doping,” *IEICE Electronics Express*, pp. 10–20130552, 2013.
- [109] M. Schroeder, “Synthesis of low-peak-factor signals and binary sequences with low autocorrelation (corresp.),” *Information Theory, IEEE Transactions on*, vol. IT-16, pp. 85 – 89, 02 1970.
- [110] J. O. Flower, G. F. Knott, and S. C. Forge, “Application of schroeder-phased harmonic signals to practical identification,” *Measurement and Control*, vol. 11, no. 2, pp. 69–73, 1978.
- [111] J. O. Flower and S. C. Forge, “Developments in frequency-response determination using schroeder-phased harmonic signals,” *Radio and Electronic Engineer*, vol. 51, pp. 226–232, May 1981.
- [112] J. J. Barnes and R. J. Lomax, “Finite-element methods in semiconductor device simulation,” *IEEE transactions on electron devices*, vol. 24, no. 8, pp. 1082–1089, 1977.
- [113] T. Adachi, A. Yoshii, and T. Sudo, “Two-dimensional semiconductor analysis using finite-element method,” *IEEE Transactions on Electron Devices*, vol. 26, no. 7, pp. 1026–1031, 1979.
- [114] R. E. Bank, D. J. Rose, and W. Fichtner, “Numerical methods for semiconductor device simulation,” *SIAM Journal on Scientific and Statistical Computing*, vol. 4, no. 3, pp. 416–435, 1983.
- [115] Silvaco, “Atlas.” Software, 2020. https://www.silvaco.com/products/tcad/device_simulation/atlas/atlas.html. Accessed 25 April 2020.
- [116] Synopsys, “Sentaurus.” Software, 2020. <https://www.synopsys.com/silicon/tcad/device-simulation/sentaurus-device.html>. Accessed 25 April 2020.
- [117] Lumerical, “Charge.” Software, 2020. <https://www.lumerical.com/products/charge/>. Accessed 25 April 2020.
- [118] D. Pérez-Galacho, D. Marris-Morini, R. Stoffer, E. Cassan, C. Baudot, T. Korthorst, F. Boeuf, and L. Vivien, “Simplified modeling and optimization of silicon modulators based on free-carrier plasma dispersion effect,” *Optics Express*, vol. 24, pp. 26332–26337, Nov 2016.

- [119] G. Rasigade, D. Marris-Morini, M. Ziebell, E. Cassan, and L. Vivien, "Analytical model for depletion-based silicon modulator simulation," *Optics Express*, vol. 19, pp. 3919–3924, Feb 2011.
- [120] D. Mishra and R. K. Sonkar, "2D modeling of silicon optical PN phase shifter," *Applied Optics*, vol. 59, pp. 998–1007, Feb 2020.
- [121] J. Van Zeghbroeck, "Principles of semiconductor devices." <http://ecee.colorado.edu/~bart/book/book/title.htm>, 2011. Accessed 12th April, 2020.
- [122] K. Iizuka, *Elements of Photonics: Vol. II for Fiber and Integrated Optics*. New York: Wiley, 2002.
- [123] D. M. Caughey and R. E. Thomas, "Carrier mobilities in silicon empirically related to doping and field," *Proceedings of the IEEE*, vol. 55, no. 12, pp. 2192–2193, 1967.
- [124] T. Liow, K. Ang, Q. Fang, J. Song, Y. Xiong, M. Yu, G. Lo, and D. Kwong, "Silicon modulators and germanium photodetectors on SOI: Monolithic integration, compatibility, and performance optimization," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 16, no. 1, pp. 307–315, 2010.
- [125] R. Ding, Y. Liu, Y. Ma, Y. Yang, Q. Li, A. E. Lim, G. Lo, K. Bergman, T. Baehr-Jones, and M. Hochberg, "High-speed silicon modulator with slow-wave electrodes and fully independent differential drive," *Journal of Lightwave Technology*, vol. 32, no. 12, pp. 2240–2247, 2014.
- [126] A. Samani, M. Chagnon, D. Patel, V. Veerasubramanian, S. Ghosh, M. Osman, Q. Zhong, and D. V. Plant, "A low-voltage 35-GHz silicon photonic modulator-enabled 112-Gb/s transmission system," *IEEE Photonics Journal*, vol. 7, no. 3, pp. 1–13, 2015.
- [127] Y. Liu, S. Dunham, T. Baehr-Jones, A. E.-J. Lim, G.-Q. Lo, and M. Hochberg, "Ultra-responsive phase shifters for depletion mode silicon modulators," *J. Lightwave Technol.*, vol. 31, pp. 3787–3793, Dec 2013.
- [128] S. Zhou, *High speed IC designs for low power short reach optical links*. PhD thesis, University College Cork, 2017.
- [129] S. Zhou, H. Wu, K. Sadeghipour, C. Scarcella, C. Eason, M. Rensing, M. Power, C. Antony, P. O'Brien, P. Townsend, and P. Ossieur, "Driver

- circuit for a PAM-4 optical transmitter using 65 nm CMOS and silicon photonic technologies,” *Electronics Letters*, vol. 52, no. 23, pp. 1939–1940, 2016.
- [130] D. Gill, W. Green, S. Assefa, J. Rosenberg, T. Barwicz, S. Shank, H. Pan, and Y. Vlasov, “A figure of merit based transmitter link penalty calculation for CMOS-compatible plasma-dispersion electro-optic mach-zehnder modulators,” *arXiv preprint arXiv:1211.2419*, 2012.
- [131] M. Power, B. Murray, S. Facchin, S. Zhou, C. Antony, C. Eason, C. Scarcella, G. Talli, P. O’Brien, P. Ossieur, and P. D. Townsend, “Low power 15 Gbaud silicon photonic link for data centres,” in *2017 European Conference on Optical Communication (ECOC)*, pp. 1–3, 2017.
- [132] E. Sonkin, D. Sadot, and G. Katz, “MZM optimization of PAM-4 transmission in data center interconnect,” *Applied Sciences*, vol. 9, no. 4, p. 637, 2019.
- [133] S. Facchin, S. Zhou, M. Power, A. Jain, C. Scarcella, C. Antony, P. Townsend, and P. Ossieur, “A 20Gbaud/s PAM-4 65nm CMOS optical receiver using 3D solenoid based bandwidth enhancement,” in *2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS)*, pp. 723–726, 2017.
- [134] B. Murray, C. Antony, G. Talli, and P. D. Townsend, “50-gbps PAM4 and electrical duobinary modulation with a lumped silicon photonic MZM,” in *45th European Conference on Optical Communication (ECOC 2019)*, pp. 1–4, 2019.
- [135] R. Marchetti, C. Lacava, A. Khokhar, X. Chen, I. Cristiani, D. J. Richardson, G. T. Reed, P. Petropoulos, and P. Minzioni, “High-efficiency grating-couplers: demonstration of a new design strategy,” *Scientific reports*, vol. 7, no. 1, pp. 1–8, 2017.
- [136] Thor Labs, “2.92mm-to-2.92mm Microwave Cables, KMM24.” Data sheet, Sept. 2017. Available at https://www.thorlabs.com/newgrouppage9.cfm?objectgroup_id=9827. Accessed 25 November 2020.
- [137] P. Torres-Ferrera, V. Ferrero, M. Valvo, and R. Gaudino, “Impact of the Overall Electrical Filter Shaping in Next-Generation 25 and 50 Gb/s PONs,” *J. Opt. Commun. Netw.*, vol. 10, pp. 493–505, May 2018.

- [138] X. Li, F. Yang, F. Zhong, Q. Deng, J. Michel, and Z. Zhou, "Single-drive high-speed lumped depletion-type modulators toward 10 fj/bit energy consumption," *Photonics Research*, vol. 5, no. 2, pp. 134–142, 2017.
- [139] IMEC, "ISIPP25G Silicon Photonics Technology Handbook," tech. rep., IMEC, 2014.
- [140] S. Caniggia and F. Maradei, *Signal Integrity and Radiated Emission of High-Speed Digital Systems*. Wiley, 2008.
- [141] D. Michelon, E. Bergeret, M. Egels, and A. Di Giacomo, "Wire-bonds used as matching inductor in RF energy harvesting applications," in *2014 10th Conference on Ph.D. Research in Microelectronics and Electronics (PRIME)*, pp. 1–4, 2014.
- [142] J. Hwang, G.-S. Jeong, W. Bae, J.-E. Park, C. S. Yoon, J. M. Yoon, J. Joo, G. Kim, and D.-K. Jeong, "A 32 Gb/s, 201 mW, MZM/EAM Cascode Push–Pull CML Driver in 65 nm CMOS," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 65, no. 4, pp. 436–440, 2018.
- [143] M. Webster, K. Lakshmikumar, C. Appel, C. Muzio, B. Dama, and K. Shastri., "Low-Power MOS-Capacitor Based Silicon Photonic Modulators and CMOS Drivers," in *Optical Fiber Communication Conference*, p. W4H.3, Optica Publishing Group, 2015.
- [144] A. Zandieh, P. Schvan, and S. P. Voinigescu, "Linear Large-Swing Push–Pull SiGe BiCMOS Drivers for Silicon Photonics Modulators," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 12, pp. 5355–5366, 2017.
- [145] S. Wolf, H. Zwickel, W. Hartmann, M. Lauermann, Y. Kutuvantavida, C. Kieninger, L. Altenhain, R. Schmid, J. Luo, A. K.-Y. Jen, S. Randel, W. Freude, and C. Koos, "Silicon-Organic Hybrid (SOH) Mach-Zehnder Modulators for 100 Gbit/s on-off Keying," *Scientific Reports*, vol. 8, no. 1, p. 2598, 2018.
- [146] C. Kieninger, C. Füllner, H. Zwickel, Y. Kutuvantavida, J. N. Kemal, C. Eschenbaum, D. L. Elder, L. R. Dalton, W. Freude, S. Randel, and C. Koos, "Silicon-organic hybrid (soh) Mach-Zehnder modulators for 100 GBd PAM4 signaling with sub-1dB phase-shifter loss," *Optics Express*, vol. 28, pp. 24693–24707, Aug 2020.



"A THESIS IS JUST LIKE CHARGING A CAPACITOR. REALLY..."
...at some point in time, we agree that it's *close enough*.