



Article

Variation of Signal Reflection on Electrodes of Silicon Mach-Zehnder Modulators: Influence of Nanoscale Variation and Mitigation Strategies

Zhaobang Zeng^{1,2,3} , Ding Ding^{1,2,3}, Qianyi Gao^{1,2,3}, Nan Yang^{1,2,3}, Peiyan Zhao^{1,2,3} and Wei Jiang^{1,2,3,*}

- ¹ College of Engineering and Applied Sciences, Nanjing University, Nanjing 210093, China; mg1534001@smail.nju.edu.cn (Z.Z.); mg1834003@smail.nju.edu.cn (D.D.); dz1934003@smail.nju.edu.cn (Q.G.); mg1634018@smail.nju.edu.cn (N.Y.); mg1434006@smail.nju.edu.cn (P.Z.)
- ² Key Laboratory of Intelligent Optical Sensing and Manipulation, Nanjing University, Ministry of Education, Nanjing 210093, China
- ³ National Laboratory of Solid State Microstructures, Nanjing 210093, China
- * Correspondence: weijiang@nju.edu.cn

Abstract: Driving signal reflection on traveling wave electrodes (TWEs) is a critical issue in Mach-Zehnder modulators. Fabrication variation often causes a random variation in the electrode impedance and the signal reflection, which induces modulation characteristics variation. The variation of reflection could be intertwined with the variation of other electrode characteristics, such as microwave signal attenuation, resulting in complexity. Here, we characterize the (partial) correlation coefficients between the reflection and modulation characteristics at different bit rates. Decreasing correlation at higher bit rates is observed. Device physics analysis shows how the observed variation can be related to nanoscale variation of material properties, particularly in the embedded diode responsible for electro-optic modulation. We develop a detailed theory to analyze two variation modes of the diode (P-i-N diode or overlapping P/N regions), which reveal insight beyond simplistic diode models. Microwave signal attenuation tends to reduce the correlation with on-electrode reflection, particularly at high bit rates. The theory shows the relative importance of conductor-induced attenuation and “dielectric”-induced attenuation, with different dependence on the frequency and fabrication variation. Strategies on how to mitigate the effect of variation for better fabrication tolerance are discussed by considering three key factors: pre-shift in structural design, bias condition, and fabrication control accuracy.



Citation: Zeng, Z.; Ding, D.; Gao, Q.; Yang, N.; Zhao, P.; Jiang, W. Variation of Signal Reflection on Electrodes of Silicon Mach-Zehnder Modulators: Influence of Nanoscale Variation and Mitigation Strategies. *Nanomaterials* **2021**, *11*, 499. <https://doi.org/10.3390/nano11020499>

Academic Editors: Andrey B. Evlyukhin and Heping Zeng
Received: 11 January 2021
Accepted: 9 February 2021
Published: 16 February 2021

Keywords: optical interconnects; optical modulation; silicon modulators; fabrication variation

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Silicon-based optoelectronic devices, which are compatible with complementary metal oxide semiconductor processes, are becoming increasingly attractive [1,2], because they have the potential to provide high-bandwidth communications with low cost and low power consumption [3]. Silicon optical modulators are key components of silicon photonics [4,5]. Mach-Zehnder modulators (MZMs) with traveling wave electrodes (TWEs) have received significant attention, owing to their broad bandwidth and reliable operation [6–16]. In reality, fabrication processes often introduce substantial structure variations that are difficult to control. Indeed, fabrication variation is an important issue that needs to be investigated for the practical use of silicon photonics. A number of studies have considered the effect of the fabrication variation. Fabrication variation of passive devices has been extensively studied [17–20]. Due to its complexity, the variation of TWE modulators performance has seldom been analyzed in depth.

For TWE MZMs, the impedances of the electrodes and their terminations often vary in fabrication processes, resulting in a significant variation of optical modulation characteristics. In our previous work [21], we analyzed the correlation between the reflection of

the driving signal and the characteristics of the output signal at a single bit rate. However, other key characteristics of TWEs, such as the attenuation of driving signal, could be intertwined with the impedance variation, causing difficulties in differentiating their individual effects. To fully understand such complex situations, further experiments and judicious quantitative analysis of low-level physical variation scenarios are needed. Note that the performance variation of modulators due to fabrication variation is crucial to the device yield, and hence it directly affects the actual device cost. As silicon modulators are key active devices in many applications, this problem can be an important factor in achieving the real promise of low cost for silicon photonics.

Here, we design a modulator with longer electrodes and investigate the correlation trend with increasing bit rates. The bit-rate dependence provides a new dimension to analyze the correlation behavior. The (partial) correlation coefficients between the driving signal reflection (S_{11}) and modulation characteristic parameters, such as the bit error rate (BER), signal-to-noise ratio (SNR), extinction ratio (ER), and jitter, are characterized. The results show that with increasing modulation bit rate, the (partial) correlation coefficients decrease gradually. To account for the observed behavior, we develop a dedicated theory to analyze the variation of key device structures, particularly the TWE and the embedded diode with nanoscale features. Fabrication variation can produce unintended variation of material properties (e.g., spatial variation of dopant concentrations) at the scale of 25~100 nm, which can substantially affect the total impedance of the electrode and cause reflection variation. The implications of the results in improving the design and fabrication of silicon MZMs are discussed.

2. Structures and Methods

Figure 1a illustrates a schematic of the modulator. Our devices were fabricated on a silicon-on-insulator wafer in a foundry with a top silicon layer thickness of 0.22 μm and a buried oxide thickness of 2 μm . The silicon ridge waveguide is 0.5 μm in width and 220 nm in height, with a 0.09 μm thick slab. The modulator was based on an asymmetric MZI structure and consisted of two phase shifters with different lengths of 20 μm between the arms, resulting in a free spectral range of ~ 30 nm. The light was coupled into and out of the chip by two grating couplers. Two multimode interference couplers (MMIs) were used as input/output 3 dB couplers. The modulation of MZMs was based on the carrier depletion.

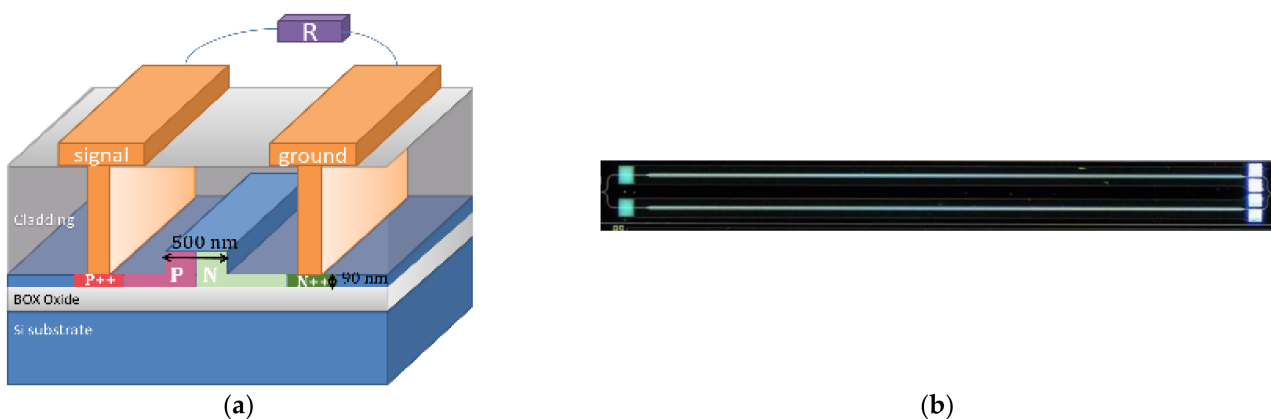


Figure 1. (a) Schematic of essential device structure and (b) optical microscopic image of a fabricated device.

A PN junction was designed with the junction interface at the center of the waveguide, and the P and N doping regions were doped to $N_A = 3.5 \times 10^{17} \text{ cm}^{-3}$ and $N_D = 2.3 \times 10^{17} \text{ cm}^{-3}$. The P and N slab regions both had a width of 1 μm . As we shall see, nanoscale variation of material doping concentration near the junction interface due to fabrication variation could play a key role in the device performance variation and will be one point of focus in this work. The P++- and N++-doped regions were 1 μm away from the edge of the ridge for ohmic contact. The position of these doped regions was designed to avoid

guided mode propagation loss. The coplanar waveguide electrode was used with a pattern of ground signal ground signal (GSGS). The characteristic impedance of our device was designed to be around 50 Ohm, assisted by the coplanar waveguide design principle and finite element simulation [9,21,22]. The G and S electrodes made out of aluminum both had a width of 60 μm and a thickness of 0.75 μm , and the gap between them was 5.5 μm . A matched resistor was integrated onto the chip to reduce the reflection of the RF signal at the output.

Figure 1b shows an optical microscope image of the device and the length of the phase shifter is 4 mm. As we shall see, such a relatively long electrode makes it easier to observe a change in the reflection-induced effect at different bit rates.

The devices were characterized experimentally. The RF driving signal reflection was characterized by a Keysight vector network analyzer (VNA) with a proper calibration process to remove the contributions of the cable and the probe. The modulation experiment was performed with the setup shown in Figure 2. A bit error rate tester (BERT) was used to provide different bit rates of the pseudo random binary sequence (PRBS) signal with a pattern length of $2^{31}-1$. The driving voltage swing was 3.2 V and the dc bias voltage is -1.7 V. The signal was applied on the MZM through a 40 GHz GS probe. The light wave from a tunable laser was controlled by a polarization controller that then coupled into the device. The typical insertion loss of the modulator itself was ~ 5 dB, which includes the losses in MMIs and waveguide propagation loss in the phase shifters. The modulated optical signal, amplified by an erbium-doped fiber amplifier (EDFA), was fed into a commercial 50 GHz photodetector. The signal was then sent back to the BERT to test the BER or was captured by an Agilent sampling oscilloscope to measure the performance parameter. All devices were set to operate at the quadrature point and the power of the receiving signal was controlled (via setting EDFA power) to be fixed at ~ 9 dBm before entering the photodetector.

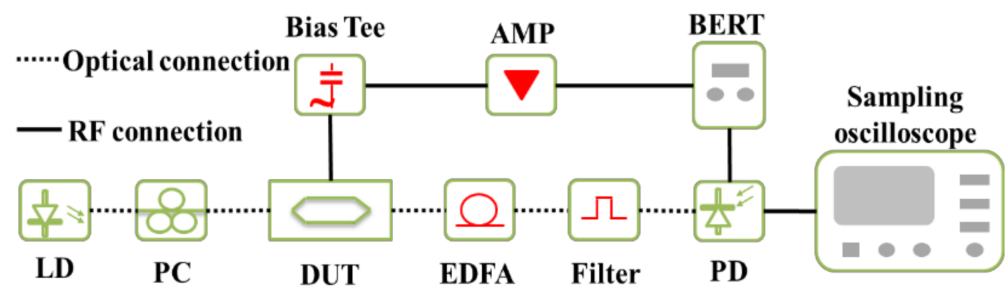


Figure 2. Experimental setup of measurement system used to characterize a Mach-Zehnder modulator (MZM).

3. Results

First, we characterize the driving signal reflection on the traveling wave electrode and the modulated signal output. The S_{11} response versus the frequency of the MZMs is measured under the 1.7 V reverse bias voltages (same voltage as in high-speed modulation experiments shown later). The representative S_{11} results are shown in Figure 3. The test data are measured from a number of devices (same design) fabricated in one batch. The S_{11} results indicate that all devices have varying degrees of mismatch between the traveling wave electrode and the terminal resistor due to fabrication variation.

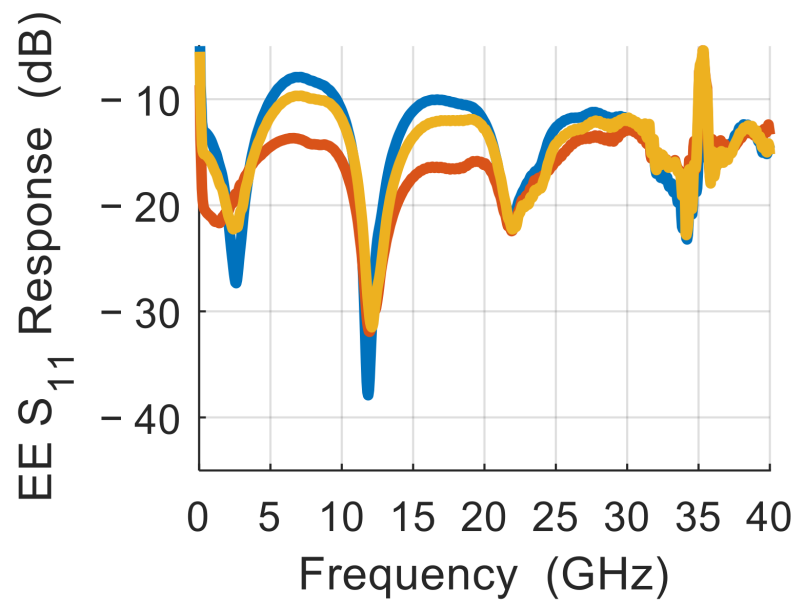


Figure 3. Representative S_{11} response (best: red; intermediate: yellow; worst: blue).

Subsequently, we characterize the modulation performance of the devices at different bit rates. Figure 4 presents the results for the (a) 22, (b) 25 and (c) 28 Gb/s eye diagrams of the representative devices. It is clear that the modulation characteristics vary substantially.

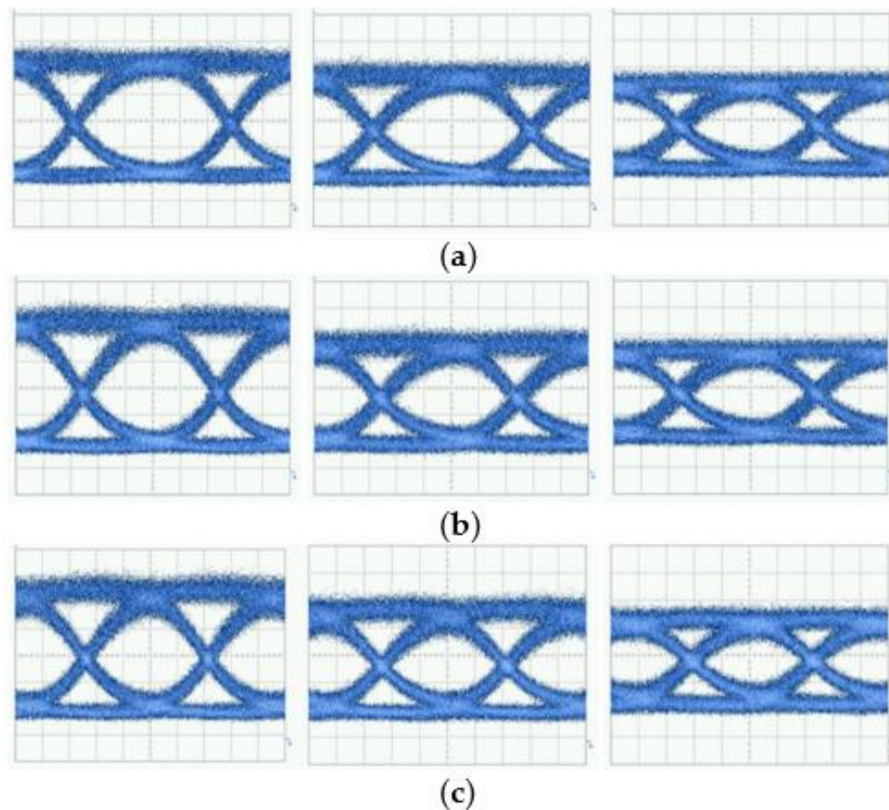


Figure 4. Representative eye diagrams at (a) 22, (b) 25, and (c) 28 Gb/s (from left to right: best, intermediate, worst). Unit for the horizontal axis: 8 ps/div; unit for the vertical axis: 20 mV/div.

The relationship between the average S_{11} response ($\langle S_{11} \rangle$) and the performance parameters in two representative frequency ranges is shown in Figure 5a,b. The frequency range for averaging S_{11} in Figure 5a is roughly 0.75 times the data rate, which contains most of the signal power [23]. The frequency range for averaging S_{11} in Figure 5b is equal to the data rate. In both figures, the relation of BER vs. $\langle S_{11} \rangle$ is almost a straight line at 22 Gb/s. As the data rate increases, some deviation from the straight line becomes increasingly obvious. Similar behaviors happen to the SNR vs. $\langle S_{11} \rangle$, although at 22 Gb/s, the deviation is already quite visible. Generally, the relationships between other parameters of the modulated signal and $\langle S_{11} \rangle$ have similar trends as the data rate increase, although the degrees of irregularity of data points may vary in each case. For example, it can be difficult to ascertain a clear trend between S_{11} and root-mean-square (RMS) jitter and peak-to-peak (PP) jitter at 28 Gb/s. Comparing Figure 5a,b, the trend of changing deviation with the data rate (from left to right) is similar. Generally, when the averaging frequency range contains most of the signal power, this trend does not change. Hence, we will use a fixed frequency range in subsequent analysis.

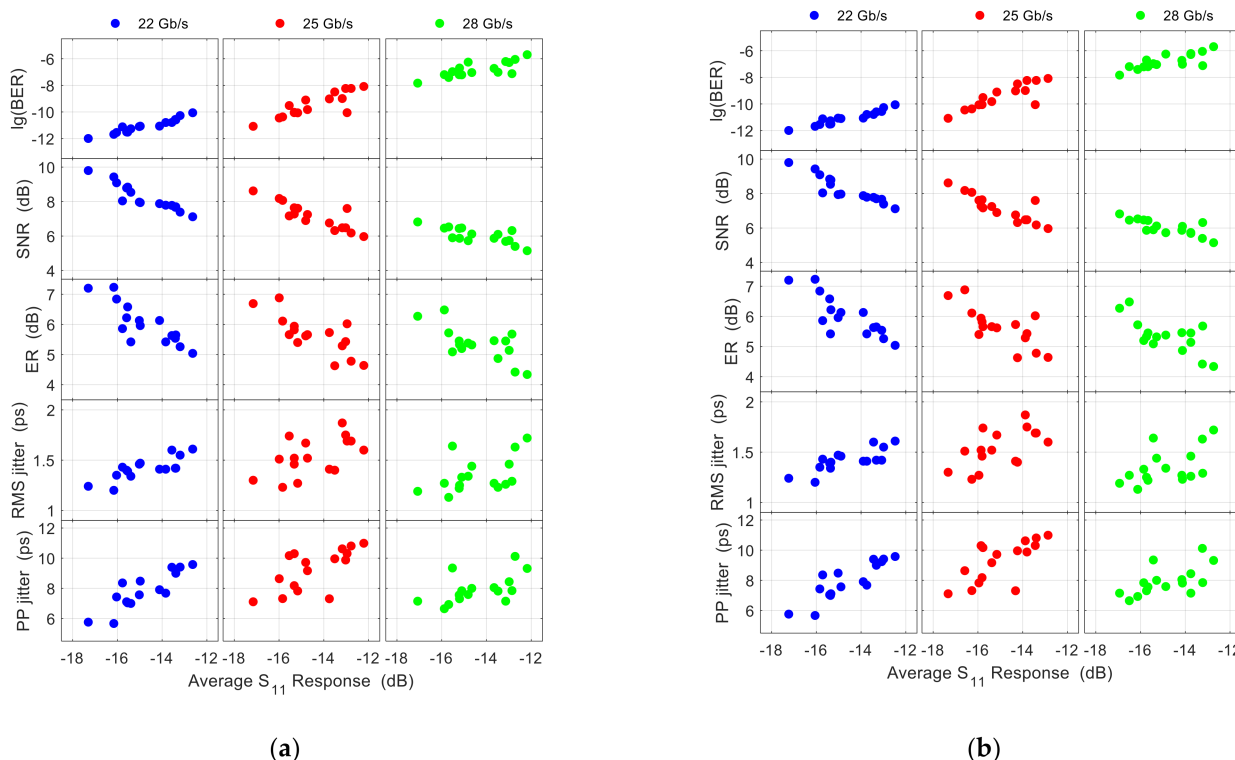


Figure 5. Relations between the bit error rate (BER), signal-to-noise ratio (SNR), extinction ratio (ER), root-mean-square (RMS) jitter, peak-to-peak (PP) jitter and the average of the S_{11} response over the frequency range of (a) 0.75 times the data rate and (b) equal to the data rate.

To quantify the relationship between $\langle S_{11} \rangle$ and performance parameters, we calculate the Pearson correlation coefficient $\rho(\langle S_{11} \rangle, x)$, where $x = \text{BER, SNR, ER, RMS jitter or PP jitter}$:

$$\rho(y, x) = \frac{\sum_i (y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum_i (y_i - \bar{y})^2 \sum_i (x_i - \bar{x})^2}} \quad (1)$$

where i is the index of the devices. Note that the amplitude of the modulated signal ΔP (the difference between states “1” and “0”) varies with the device.

To remove the influence of the amplitude variation, we can calculate the partial correlation coefficient [20]. The partial correlation coefficient between two parameters,

y and x , while holding the third parameter z ($=\Delta P$ here) constant was calculated via the standard form [24]:

$$p(y, x; z) \equiv \frac{[\rho(y, x) - \rho(y, z)\rho(x, z)]}{\{[1 - \rho(y, z)][1 - \rho(x, z)]\}^{\frac{1}{2}}} \quad (2)$$

The correlation coefficients $\rho(\langle S_{11} \rangle, \Delta P)$, $\rho(x, \Delta P)$ in the formula of $p(\langle S_{11} \rangle, x; \Delta P)$ can be readily calculated.

To illustrate how the correlation changes with data rate, we plot the (partial) correlation coefficients between $\langle S_{11} \rangle$ and the strongly correlated performance parameters (BER, SNR, and ER) in Figure 6. Here, S_{11} is averaged over the full frequency range ($f_{max} = 40$ GHz) of our test system. Choosing another averaging frequency range (as long as it contains most of the signal power) will result in a similar trend. It is evident that both the correlation coefficient and the partial correlation coefficient decrease as the modulation rate increases. Generally, Figures 5 and 6 show that BER, SNR, and ER are more affected by the RF signal reflection, whereas the signal timing (jitters) is less affected, which is understandable as the driving signal reflection directly affects the signal amplitude and has only a secondary influence on the signal timing through non-steep transition edges [21]. Note that the reflected signal is superposed on the original driving signal. As the reflected signal often carries the information from the previous bit(s), it may randomly increase or decrease the amplitude of the current bit of the driving signal. Hence the reflection generates noise, which explains its correlation with SNR. The BER is often directly related to the SNR (via well-known formulas, see Ref. [25] for example). The ER variation originates largely from the nonlinear relation between driving voltage and modulated output signal [21].

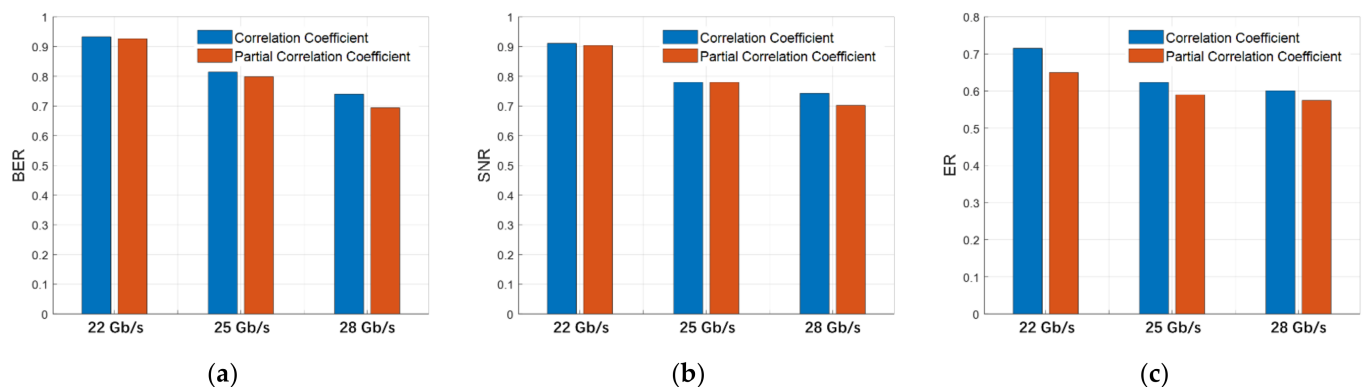


Figure 6. Absolute value of the correlation coefficients and partial correlation coefficients of average S_{11} with performance parameters: (a) BER, (b) SNR, and (c) ER.

To understand the decreasing correlation with increasing data rate, we note that the reflected signal on the electrode is superposed on the original driving signal, which can distort the original signal, thus causing different effects on the various characteristics of the modulated signal in these devices. With increasing modulation data rate, the fraction of high-frequency content in the driving signal will increase. When the signal reaches the TWE end and is reflected, the reflected signal tends to be weaker because more high-frequency content experiences more attenuation, and the reflected signal will be attenuated more along the reverse propagation path and thus produces fewer effects. Although the forward propagating driving signal is also attenuated, its attenuation is less than the reflection attenuation. For example, if the attenuation of the forward signal is 0 to u dB from the beginning to the end of the TWE, the reflected signal will be attenuated by u to $2u$ dB from the end of TWE to the beginning. Obviously, the reflected signal is attenuated more. To further illustrate this, one can imagine an extreme case that the high frequency signal is attenuated to almost zero at the TWE end, the reflection would have virtually no effect on the device performance. In addition, the reflected RF signal is counter-propagating on the electrode. As such, theory shows that the modulation effect due to reflected RF

signal typically decreases with frequency (with a sinc function type of behavior in some simplified models) [26], which suggests a reduced modulation bandwidth of the reflected signal compared to that of the original driving signal. This is another reason that the effect of reflection is less for faster bit rates. Therefore, the relative contribution of reflection might become less at high frequencies (or high data rates), which suggests lower correlation.

4. Discussion

To fully understand where the variation of reflection comes from, we need delve into the underlying structure and investigate how the variation of TWE structure parameters influences the reflection. The output power for MZMs depends on the electro-optic phase shift $\Delta\varphi$ as $P = P_0 + P_1 \cos(\Delta\varphi - \Delta\varphi_0)$, where P_0 , P_1 and $\Delta\varphi_0$ are constants. The phase shift $\Delta\varphi$ is proportional to the refractive index change Δn , which depends on the effective voltage on the TWEs. The effective voltage on the TWEs is affected by the reflected signal caused by the terminal mismatch, resulting in the difference of the modulation performance parameters between the devices, as shown in Figure 4. As shown in Figure 2, the peak value of the S_{11} difference between the best and worst devices is over 5 dB. Note that the terminal mismatch is caused by the variation in electrode impedance and terminal resistance. Based on the DC measurement, the resistance of the terminating resistor varies between 48 and 53 Ω . The electrode impedance is related to a large number of large structural features of the electrode (including the gap and thickness) and small features within the optical waveguide (such as the diode with nanoscale junction width), as shown in Figure 1a. Note that the variation of structural dimensions is typically at the scale of 25~100 nm for some key features. In addition, the degree of influence that each parameter exerts on its respective structure may differ. However, with the presence of the on-chip termination, it is difficult to directly measure the frequency-dependent total impedance of each individual device.

In order to find how the variation of the impedance is related to the variation of the low-level structure parameters, a widely used modeling method [8,14,26,27] is employed here. In this method, the PN loaded characteristic impedance Z_0 and propagation constant γ can be expressed as:

$$Z_0 = \sqrt{\frac{R_{tl} + j\omega L_{tl}}{(G_{tl} + R_{tpn}^{-1}) + j\omega(C_{tl} + C_{tpn})}} \quad (3)$$

$$\gamma = \alpha + j\beta = \sqrt{(R_{tl} + j\omega L_{tl}) \left([G_{tl} + R_{tpn}^{-1}] + j\omega [C_{tl} + C_{tpn}] \right)} \quad (4)$$

where R_{tl} , L_{tl} , C_{tl} and G_{tl} are the resistance, inductance, capacitance, and leakage conductance of the transmission line per unit length, respectively, α is the amplitude attenuation constant, β is the phase constant, and R_{tpn} and C_{tpn} are the parallel equivalent values of the diode's resistance and capacitance, respectively. Finite-element method (FEM) simulations in the 2D cross-section of the modulator are performed to obtain the transmission line parameters of the electrodes. The P-N diode's characteristics are largely governed by a one-dimensional differential equation along the horizontal axis, which can be readily solved with well-established analytic formula [28].

Considering that the width variation of the electrode metal dimensions is about 2%, we simulated the influence of the change in the gap between the metal electrode. In Figure 7, the blue solid line is the simulation result obtained at the initial design value. Figure 7a,b shows the change in the total impedance of the PN junction loaded transmission line and the attenuation constant when the electrode gap is changed. When the frequency is lower than 10 GHz, there is almost no difference in the impedance and the attenuation constant of different gap variations. In the full test frequency range, when the gap change is $\pm 0.3 \mu\text{m}$, the total impedance changes by about $\pm 0.3 \Omega$. Note that for such a gap ($5.5 \mu\text{m}$), the range of width variation (maximum minus minimum) is generally substantially less than

10% [29]. The variation in the gap of the electrode caused by the fabrication variation has minimal effect on the variation of the transmission loss of the TWEs. Note that because the electrode width is one order of magnitude larger than the gap width, the electrode width variation (while the gap width is fixed) produces a negligible effect compared to the gap width variation. In addition, we also simulate the influence of the change in thickness of the TWEs. Typically, the range of metal film thickness variation should be less than 10% [29]. Figure 7c,d show the simulation results for the thickness varying by $\pm 0.05 \mu\text{m}$. As the thickness of the electrode varies by $0.05 \mu\text{m}$, the impedance of the PN junction loaded transmission line changes is around 0.1Ω in the test frequency range. The variation of the microwave attenuation influenced by the variation of the thickness is around -1% . In general, the variation of the electrode's structural characteristics has a very small effect on the transmission line impedance ($\pm 1\%$), which cannot account for the observed over 5 dB of variation of RF signal reflection.

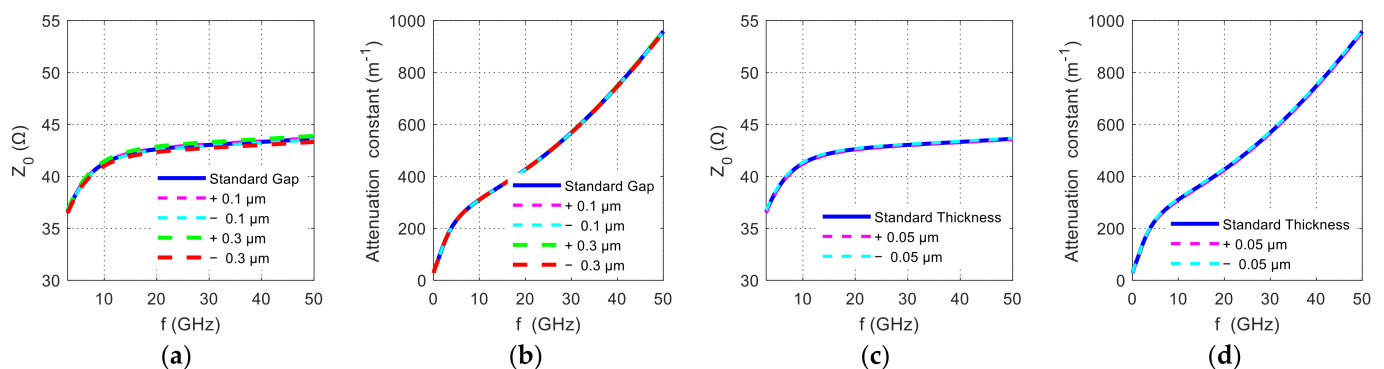


Figure 7. Simulated variation of the TWE with the P-N junction loaded: (a) total impedance and (b) total power attenuation constant varying with the gap between G and S; (c) total impedance and (d) total power attenuation constant varying with the electrode thickness.

Compared to the features of the electrodes, the edges of the P and N regions in submicron optical waveguides are more likely to shift due to multiple processing steps. For example, the edge of each region may obviously shift due to misalignment during the lithography. Furthermore, the width of the P or N region may expand or shrink during the lithography and etching steps, which may also cause the edge of each region to shift. The edge shift of the P and N doping regions causes unintended spatial variation of materials property (i.e., dopant concentration). Although the linewidth variation and misalignment can be controlled to the order of 25~100 nm, this may still have a relatively large effect on the capacitance of the PN junction because the junction width itself is of the order of 100 nm. As shown in Figure 8, fabrication variation may cause two variation scenarios of the diode. The first type is the diode becoming a P-i-N diode (i.e., with a nano-gap between the P and N regions), and the second type causes the diode to have a nanoscale overlapping region. Either case can cause substantial junction capacitance variation.

For the first type, the junction capacitance can be modeled by a relatively simple formula [28]:

$$C_{pin} = C_{pn} \frac{1}{\sqrt{1 + W_i^2 / W_{pn}^2}} \quad (5)$$

where W_i is the nano-gap width between the P and N regions and W_{pn} and C_{pn} are the junction width and capacitance of a corresponding PN diode, respectively. Figure 9a shows the junction capacitance variation with bias voltage for different gap widths. At the same bias voltage, different gap widths give rise to different junction capacitance drop rates (e.g., when the reverse bias voltage is 1 V, the junction capacitance drops by $\sim 5 \text{ fF/mm}$ when the gap is increased from 0 to 30 nm and by $\sim 15 \text{ fF/mm}$ when the gap is increased from 90 to

120 nm). In addition, when the bias voltage is smaller, the junction capacitance changes at a faster rate with the change in gap width.

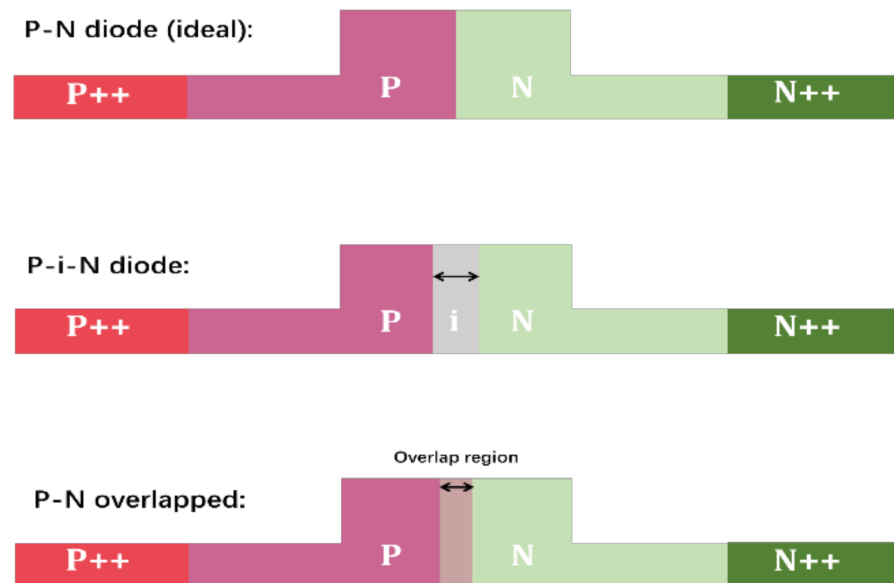


Figure 8. Two fabrication variation scenarios of the diode.

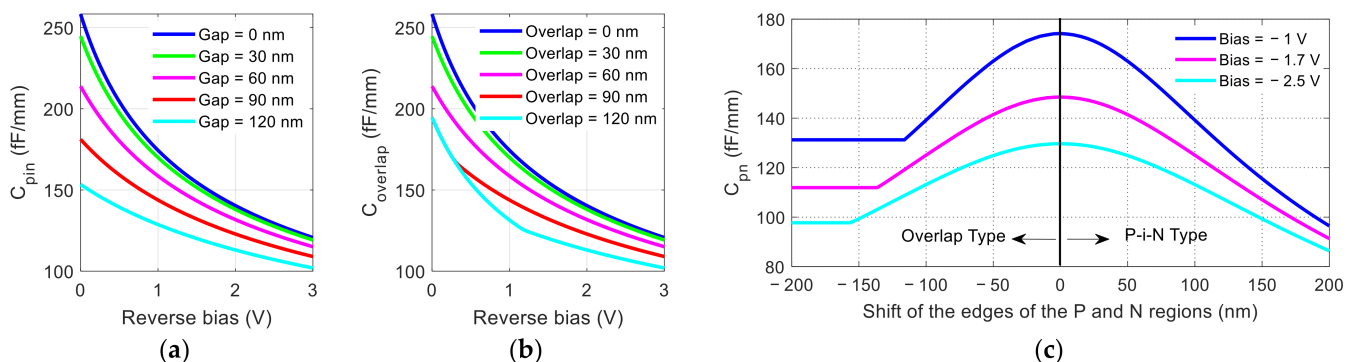


Figure 9. Calculated variation of junction capacitance with bias voltage for (a) P-i-N type and (b) overlap type; and (c) diode under different biases for both types.

For the second type, the overlapping region will be P doped with a net concentration of $N_{A^-} = N_A - N_D = 1.2 \times 10^{17} \text{ cm}^{-3}$. Here, two cases must be considered. First, if the overlap region is wide and the bias voltage is small, the overlap region may not be fully depleted. In this case, the capacitance can be readily calculated for a PN junction where the P side is doped to a concentration of N_{A^-} . Second, if the overlap region is narrow and the bias voltage is sufficient to deplete the overlap region to reach the fully doped P region, the capacitance of such a junction is more complicated. For this type, detailed derivation in the Appendix A shows that the junction capacitance is:

$$C_{ov} = \frac{\epsilon_{Si}}{\sqrt{\gamma_A^2 w_{ov}^2 + \frac{2\epsilon_{Si}}{qN_D} \gamma_A (\psi_0 - V)} + \sqrt{\gamma_D^2 w_{ov}^2 + \frac{2\epsilon_{Si}}{qN_A} \gamma_D (\psi_0 - V)}} \quad (6)$$

where q is the electron charge, ϵ_{Si} is the dielectric constant of silicon, w_{ov} is the overlap width, ψ_0 is the built-in potential, V is the bias. The definition of γ_A and γ_D is given in Equation (A8) in Appendix A. Figure 9b shows the variation of junction capacitance with bias voltage for different overlap widths. Different from the P-i-N diode scenario, when the

overlap area is large, the curves may merge (for example, when the reverse bias was less than 0.4 V, the junction capacitance curves corresponding to an overlap of 90 nm and an overlap of 120 nm have the same value). It can be shown that this happens due to the first case where the small bias is not sufficient to deplete a 90 nm wide overlap region. Hence, when the overlap width is increased to 120 nm, the depletion region is not affected, and the capacitance remains the same.

Figure 9c shows the junction capacitance changes in the two scenarios of variations under different bias voltages (positive or negative shift of the edges of the P and N regions corresponds to the P-i-N or overlap scenario, respectively). Both variation scenarios of the diode lead to a reduction in PN junction capacitance. Furthermore, for the P-i-N case, the capacitance continues to decrease as the i region increases, while for the overlap scenario, when the overlap region is very large, the capacitance reaches a stable value (indicating that the voltage is insufficient to deplete the overlap region).

Figure 10a shows the FEM simulation result of the total impedance changes corresponding to the C_{pn} changes. If the individual shift of the P and N edges is ~ 100 nm (hence the worst relative shift under opposite shift directions is ~ 200 nm), the total impedance of the electrode increased by 8% or more, which can produce substantial reflection variation as observed experimentally. Compared with the change in the electrode parameters, the change in the impedance caused by the change in the location of the P and N regions is more significant, which is a major factor for the reflection variation between the devices.

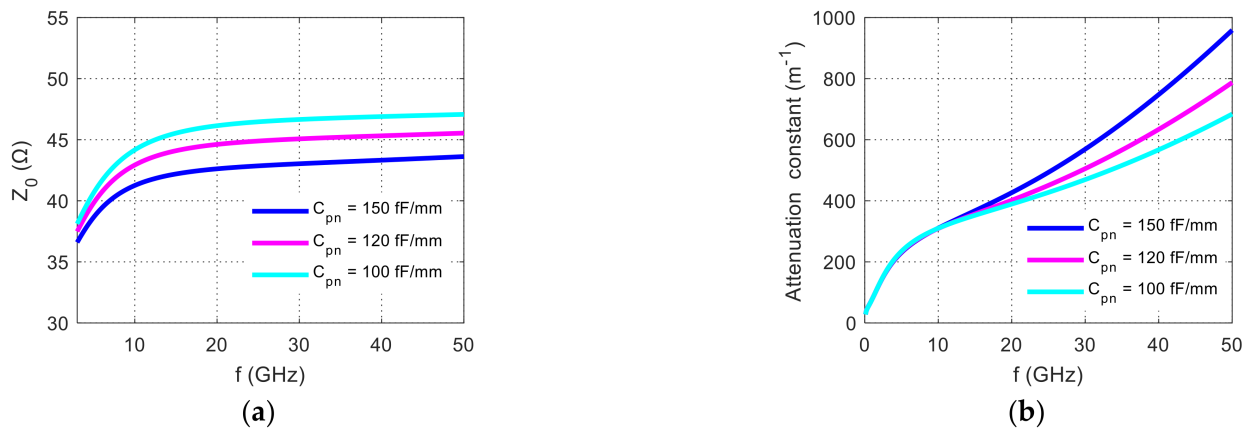


Figure 10. Simulation results of (a) total impedance and (b) total power attenuation constant varying with the junction capacitance.

In order to better understand variation of RF attenuation due to the fabrication variation of the diode, we note that the amplitude attenuation constant can be roughly approximated as [14]:

$$\alpha = \frac{1}{2} \left(R_{DC} + R_{AC} \sqrt{f} \right) \sqrt{\frac{C_{tl} + C_{pn}}{L_{tl}}} + \frac{1}{2} \left(4\pi^2 f^2 C_{pn}^2 R_{pn} \right) \sqrt{\frac{L_{tl}}{C_{tl} + C_{pn}}} \quad (7)$$

where the first term is the loaded conductor loss, the second term is the silicon “dielectric” loss, and R_{DC} and R_{AC} are some effective constants. Note that the “dielectric” loss increases substantially with the PN junction capacitance and with frequency according to Equation (7). When the junction capacitance changes, the transmission loss also changes. Here, we calculate the total attenuation constant and plot it in Figure 10b. When the frequency is low (< 15 GHz), the difference in the total microwave attenuation constant caused by the change in the C_{pn} is negligible. As the frequency increases, the variation in the total microwave attenuation constant gradually increases. This can be clearly understood through Equation (7). At low frequencies, the conductor loss ($\sim \sqrt{f}$) given by the first term dominates, whereas at high frequencies, the “dielectric” loss ($\sim f^2$) given by the second

term dominates. One readily shows that the C_{pn}^2 factor in the second term gives much higher relative variation than the $\sqrt{C_{tl} + C_{pn}}$ factor. Hence the overall loss exhibits more variation at high frequencies where the “dielectric” loss dominates. As most signal power is contained in the relatively low frequency range, the power in the frequency range above 15 GHz is less than ~15% for 22~28 Gb/s signals. Hence, the effective attenuation variation is fairly small (estimated $< \sim 1.5\%$) for the modulated output signals. Note that impedance variation $\sim 10\%$ shows up at a much lower frequency (~ 7 GHz) and continues to high frequencies, affecting around half of the signal power. Hence, the effective impedance variation is fairly substantial.

The reflection variation of the driving signal on the TWEs is one of the most common fabrication variations in silicon MZMs. In order to mitigate the modulation characteristics variation due to the on-electrode reflection variation, more emphasis should be placed on reducing diode variation. According to the above analysis, the variation of the capacitance of the PN junction is affected by the fabrication control accuracy and the bias voltage, hence proper setting of these two parameters may help mitigate the effect of reflection variation. Furthermore, it is possible to pre-shift edges of the P and N regions in design towards the goal of avoiding some worst scenarios of the relative shift. To find promising strategies, Figure 11 shows the maximum variation of the junction capacitance when the P and N region’s edges with different pre-shifts are subjected to different fabrication control accuracy (i.e., different fabrication variation). Two representative bias voltages are considered in Figure 11a,b. Overall, the left half of each curve is generally lower than the right half (divided at the zero pre-shift), which suggests that negative pre-shift or the overlap case is preferred to reduce junction capacitance variation. A typical curve can be divided into seven segments by cusp points. At each cusp point, the worst variation case swaps from the negative fabrication-induced shift to positive, or vice versa.

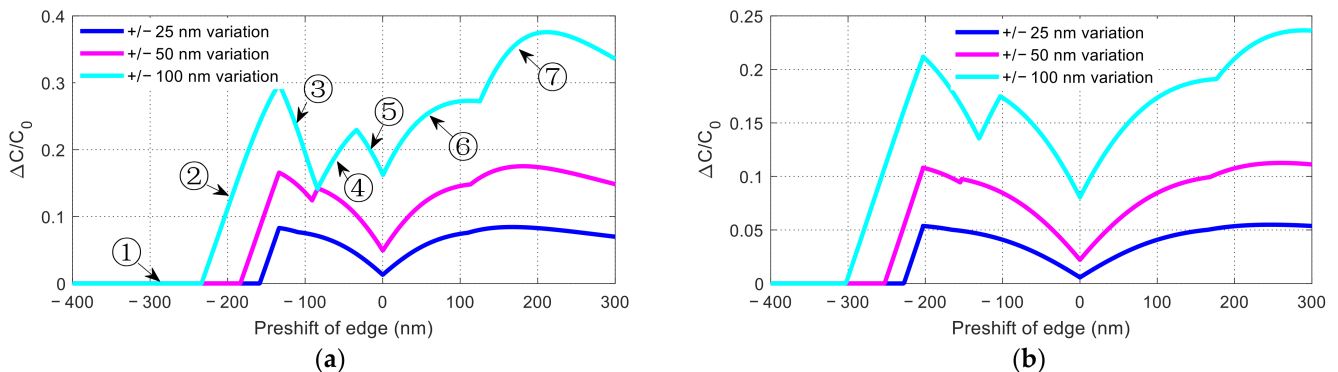


Figure 11. Mitigation of fabrication variation: Junction capacitance affected by the edge shift of the P and N regions caused by fabrication variation for different relative pre-shift of the edges. (a) under -1.7 V bias, (b) under -5 V bias. $C_0 = C_{pn}$ is the original PN junction capacitance with no overlap or gap. Positive pre-shift: P-i-N case. Negative pre-shift: overlap case.

To mitigate the capacitance variation at a bias voltage -1.7 V, for a maximum fabrication variation of ± 100 nm, a pre-shift of -80 nm between the P and N regions appears to be a fairly good setting, indicated by a local minimum on the top curve in Figure 11a. This means that a small overlap region (80 nm wide) can be designed in advance to control the manufacturing deviation. For better nanofabrication control (i.e., smaller fabrication variation) of ± 25 and ± 50 nm, zero pre-shift (i.e., no intentional shift) appears to be a better choice according to the local minimum of the respective curves, which is beyond our expectations (preliminary exploration focused on the P-i-N case only [21] is inclined to better tolerance with a sufficiently wide i region). Note that the ± 100 nm curve almost touches the ± 50 nm curve at a pre-shift of -80 nm. This is an interesting exception that occurs accidentally due to: (1) a constant capacitance for the relative shift < -130 nm ($= -80-50$ nm) and (2) the peak at zero shift (within 100 nm from the pre-shift of 80 nm), as can be inferred from Figure 9c. To mitigate the capacitance variation to nearly zero,

there is always the option of creating a large overlap region for the PN diode, as shown by the left-most segment (Segment 1) in Figure 11a. As shown in Figure 11a, the capacitance remains constant in the left-most segment. In this scenario, the carriers in the large overlap region will not be depleted at a given reverse bias and “break through” to the P doped region will not occur even with fabrication variation. As a result, the junction capacitance has the best fabrication tolerance. However, the large overlap approach may have some implications, which will be more obvious in subsequent discussion on high-bias scenarios.

As some applications may have a modulator working at a high bias voltage (e.g., [30]), we also show the capacitance variation under a larger bias voltage (-5 V) in Figure 11b. It can be seen that when the bias voltage is larger, the capacitance generally varies less for the same fabrication variation compared to Figure 11a. Intuitively, a larger bias tends to create a wider junction depletion region, which is less sensitive to P/N edge shift. Mathematically, this can be attributed to the fact that the capacitance generally contains $\sqrt{c_0(V_0 - V)}$ type of dependence on voltage (in denominator), both in the P-i-N and overlap case, as shown in Equations (5) and (6). Note that C_{pm} in Equation (5) contains a factor $\sim 1/\sqrt{c_0(V_0 - V)}$. One can readily show that such a kind of voltage dependence tends to render the capacitance insensitive at high bias. Note that at larger bias voltages, if one chooses to create a wide overlap region to achieve near zero capacitance variation, the required overlap width may be too large (e.g., 300 nm of edge pre-shift is required at a relative edge shift change of ± 100 nm at -5 V). As such, the P and N region edges may nearly touch the sidewall of the waveguide (only 100 nm margin on each side for a 500 nm wide waveguide), which may cause difficulty for carrier transport into the upper portion of the waveguide ridge. To avoid this, one may need better fabrication control (e.g., ± 50 or 25 nm, indicated by the two lower curves in Figure 11b, so that one can work at the local minimum at zero pre-shift).

In this work, we take a hybrid approach to the fabrication-induced variation of the Radio Frequency (RF) signal reflection on the modulator’s electrode. At the device level, we adopt a more experimental approach to find the correlation between variation of the modulation characteristics and that of the reflection. At the low level (structure level), we use theoretical/numeric analysis to explore the underlying variation inside the electrode (plus the diode) that can contribute to the reflection variation. Such an approach helps to overcome some difficulties at the device level and at the structure level. At the structure level, the low-level structure variation (especially, the diode junction variation) is often hard to measure experimentally and extensively. At the device level, modeling of the large-signal digital modulation for the silicon modulator is substantially more complicated, considering the nonlinear dependence between the driving signal and output modulated signal, pseudo-random bit sequence and other complications. Building a full simulation software package from low-level structure variation to digital modulation characteristics (e.g., BER, SNR) variation is beyond the scope of this work. Note that the widely used modeling method of the electrode impedance for junction-loaded modulators [8,14,26,27] is also a hybrid method that combines a direct solution of the Maxwell equation for the metal electrode plus an equivalent circuit model for the loaded diode. A hybrid approach divides a complicated problem into several parts so that each part can be solved by the most effective method, which is often useful in tackling some very complicated problems.

5. Conclusions

This work investigates the correlation between driving signal reflection variation on the electrode and modulation characteristics variation at different bit rates. Decreasing correlation with increasing bit rates is observed and is attributed to increasing high-frequency content at higher bit rates, for which reflection produces fewer effects due to increased signal attenuation and due to the smaller bandwidth of counter propagating interaction.

Low-level theoretical/numeric analysis of the underlying structures shows that the reflection variation on the electrode is less affected by the fabrication variation of the electrode itself and more affected by the variation in the embedded diode, particularly

nanoscale variation of material doping concentrations. We develop a detailed theory to analyze the two modes (P-i-N or overlap) of variation of the embedded diodes. It is shown that the capacitance variation is influenced by the fabrication control accuracy and the bias voltage.

We also provide strategies on how to reduce the effects of fabrication variation analyzed herein. When the reverse bias voltage is small, a diode with a relatively narrow overlapping P/N region (e.g., 80 nm at -1.7 V) might be a reasonably good design choice when the fabrication variation is large. With reduced fabrication variation (e.g., 25~50 nm), an “aboriginal” diode design with zero pre-shift of P/N edges appears to be reasonably good. Generally, a sufficiently large overlap region could in principle achieve zero capacitance variation, although the required overlap width may be excessively large under large bias voltages, leading to other concerns.

The methodology and results of this work may also help in studying the fabrication variation of other integrated photonic devices.

Author Contributions: N.Y. and W.J. were involved in the device design; Z.Z. and D.D. were in charge of the experimental test; Z.Z. and W.J. worked on the theory and analysis; Z.Z. wrote the paper with support from Q.G., P.Z. and W.J. All authors contributed to the technical discussions. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded in part by the National Natural Science Foundation of China under Grant 61775094, the Jiangsu Innovative/Entrepreneurial Teams, and the National Key R&D Program of China under Grants 2017YFA0303700 and 2017YFA0303704.

Data Availability Statement: The data presented in this study are available on request from the corresponding author W.J.

Acknowledgments: We thank Patrick G.-Q. Lo for helpful discussion.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A

To study the overlap scenario of a diode, we consider the following model. For the P region and P^- region side, the acceptor concentration is $N_A(x)$ is a step function comprising two segments of constant values (N_A and $N_{A^-} = N_A - N_D$). If the voltage is small, such that the overlap region is not fully depleted, this junction can be easily treated as a PN junction. However, if the overlap region is fully depleted and the depletion region extends to the P region, the simple PN junction capacitance formula no longer applies. In this case, the relative potential for the PP^- side becomes:

$$\psi_P = \frac{-q}{\epsilon_{Si}} \int_{-x_p}^0 x N_A(x) dx \quad (A1)$$

$$\psi_P = (q/\epsilon_{Si}) [N_A x_p^2/2 - N_D x_{ov}^2/2] \quad (A2)$$

where q is the electron charge, ϵ_{Si} is the dielectric constant of silicon, $-x_p$ is the location of the depletion edge in the P region, w_{ov} is the overlap width. The relative potential of the N side is:

$$\psi_N = \frac{q}{\epsilon_{Si}} \int_0^{x_n} x N_D(x) dx = \frac{q}{\epsilon_{Si}} N_D (x_n^2/2) \quad (A3)$$

where x_n is the location of depletion edge in the N region. The total potential across the junction satisfies the relationship:

$$\psi_P + \psi_N = \psi_0 - V \quad (A4)$$

where V is the bias and the built-in potential is given by:

$$\psi_0 = \frac{1}{2} \frac{kT}{q} \ln \left(\frac{N_A^2 N_D^2}{n_i^4} \right) \quad (\text{A5})$$

where k is the Boltzmann constant, T the temperature, and n_i is the intrinsic carrier concentration. In thermal equilibrium, the total negative charge per unit area in the P side and P⁻ side must be equal to the total positive charge per unit area in the N side:

$$N_A(x_p - w_{ov}) + (N_A - N_D)w_{ov} = N_D x_n \quad (\text{A6})$$

Combining (11), (12), (13), we have

$$N_D(x_n + w_{ov})^2 + N_A(x_n^2 - w_{ov}^2) = \frac{N_A(2\varepsilon_{Si}/q)(\psi_0 - V)}{N_D} \quad (\text{A7})$$

which gives x_n as an implicit function of V . In addition, according to (13), $x_p = (N_D/N_A)(x_n + w_{ov})$, and note that the depletion width is given by $W = x_p + x_n$. Through lengthy calculation, one can obtain the depletion width $W(V)$ as a function of voltage

$$W(V) = \sqrt{\gamma_A^2 w_{ov}^2 + \frac{2\varepsilon_{Si}}{qN_D} \gamma_A (\psi_0 - V)} + \sqrt{\gamma_D^2 w_{ov}^2 + \frac{2\varepsilon_{Si}}{qN_A} \gamma_D (\psi_0 - V)} \quad (\text{A8})$$

where

$$\gamma_A = \frac{N_A}{N_A + N_D}, \quad \gamma_D = \frac{N_D}{N_A + N_D} \quad (\text{A9})$$

The result of the depletion layer capacitance is $C(V) = \varepsilon_{Si}/W(V)$.

References

- Soref, R. The past, present, and future of silicon photonics. *IEEE J. Sel. Top. Quantum Electron.* **2006**, *12*, 1678–1687. [[CrossRef](#)]
- Jalali, B.; Fathpour, S. Silicon Photonics. *J. Lightwave Technol.* **2006**, *24*, 4600–4615. [[CrossRef](#)]
- Miller, D.A.B. Optical interconnects to electronic chips. *Appl. Opt.* **2010**, *49*, F59–F70. [[CrossRef](#)] [[PubMed](#)]
- Liow, T.Y.; Ang, K.W.; Fang, Q.; Song, J.F.; Xiong, Y.Z.; Yu, M.B.; Lo, G.Q.; Kwong, D.L. Silicon Modulators and Germanium Photodetectors on SOI: Monolithic Integration, Compatibility, and Performance Optimization. *IEEE J. Sel. Top. Quantum Electron.* **2010**, *16*, 307–315.
- Miller, D.A.B. Device Requirements for Optical Interconnects to Silicon Chips. *Proc. IEEE* **2009**, *97*, 1166–1185. [[CrossRef](#)]
- Reed, G.T.; Mashanovich, G.; Gardes, F.Y.; Thomson, D.J. Silicon optical modulators. *Nat. Photonics* **2010**, *4*, 518–526. [[CrossRef](#)]
- Dong, P.; Chen, L.; Chen, Y.K. High-speed low-voltage single-drive push-pull silicon Mach-Zehnder modulators. *Opt. Express* **2012**, *20*, 6163–6169. [[CrossRef](#)]
- Ding, J.F.; Chen, H.T.; Yang, L.; Zhang, L.; Ji, R.Q.; Tian, Y.H.; Zhu, W.W.; Lu, Y.Y.; Zhou, P.; Min, R. Low-voltage, high-extinction-ratio, Mach-Zehnder silicon optical modulator for CMOS-compatible integration. *Opt. Express* **2012**, *20*, 3209–3218. [[PubMed](#)]
- Wang, J.; Qiu, C.; Li, H.; Ling, W.; Li, L.; Pang, A.; Sheng, Z.; Wu, A.M.; Wang, X.; Zou, S.C.; et al. Optimization and Demonstration of a Large-bandwidth Carrier-depletion Silicon Optical Modulator. *J. Lightwave Technol.* **2013**, *31*, 4119–4125. [[CrossRef](#)]
- Gill, D.M.; Xiong, C.; Proesel, J.E.; Rosenberg, J.C.; Orcutt, J.; Khater, M.; Ellis-Monaghan, J.; Stricker, A.; Kiewra, E.; Martin, Y.; et al. Demonstration of Error-Free 32-Gb/s Operation from Monolithic CMOS Nanophotonic Transmitters. *IEEE Photonics Technol. Lett.* **2016**, *28*, 1410–1413. [[CrossRef](#)]
- Xiao, X.; Li, M.F.; Wang, L.; Chen, D.G.; Yang, Q.; Yu, S.H. High speed silicon photonic modulators. In Proceedings of the 2017 Optical Fiber Communications Conference and Exhibition (OFC), Los Angeles, CA, USA, 19–23 March 2017.
- Li, M.F.; Wang, L.; Li, X.; Xiao, X.; Yu, S.H. Silicon intensity Mach-Zehnder modulator for single lane 100 Gb/s applications. *Photonics Res.* **2018**, *6*, 109–116. [[CrossRef](#)]
- Jacques, M.; Xing, Z.; Samani, A.; El-Fiky, E.; Li, X.; Xiang, M.; Lessard, S.; Plant, D.V. 240 Gbit/s Silicon Photonic Mach-Zehnder Modulator Enabled by Two 2.3-Vpp Drivers. *J. Lightwave Technol.* **2020**, *38*, 2877–2885. [[CrossRef](#)]
- Patel, D.; Ghosh, S.; Chagnon, M.; Samani, A.; Veerasubramanian, V.; Osman, M.; Plant, D.V. Design, analysis, and transmission system performance of a 41 GHz silicon photonic modulator. *Opt. Express* **2015**, *23*, 14263–14287. [[CrossRef](#)] [[PubMed](#)]
- Wang, J.; Zhou, J.; Zhu, L.; Zhang, Q. Frequency- and Time-Domain Modeling and Characterization of PN Phase Shifters in All-Silicon Carrier-Depletion Modulators. *J. Lightwave Technol.* **2020**, *38*, 4462–4469. [[CrossRef](#)]
- Sepehrian, H.; Yekani, A.; Rusch, L.A.; Shi, W. CMOS-Photonics Codesign of an Integrated DAC-Less PAM-4 Silicon Photonic Transmitter. *IEEE Trans. Circuits Syst. I Regul. Pap.* **2016**, *63*, 2158–2168. [[CrossRef](#)]

17. Chen, X.; Mohamed, M.; Li, Z.; Shang, L.; Mickelson, A.R. Process variation in silicon photonic devices. *Appl. Opt.* **2013**, *52*, 7638–7647. [[CrossRef](#)] [[PubMed](#)]
18. Xing, Y.; Spina, D.; Li, A.; Dhaene, T.; Bogaerts, W. Stochastic collocation for device-level variability analysis in integrated photonics. *Photonics Res.* **2016**, *4*, 93–100. [[CrossRef](#)]
19. Lu, Z.Q.; Jhoja, J.; Klein, J.; Wang, X.; Liu, A.; Flueckiger, J.; Pond, J.; Chrostowski, L. Performance prediction for silicon photonics integrated circuits with layout-dependent correlated manufacturing variability. *Opt. Express* **2017**, *25*, 9712–9733. [[CrossRef](#)]
20. Boeuf, F.; Cremer, S.; Temporiti, E.; Fere, M.; Shaw, M.; Baudot, C.; Vulliet, N.; Pinguet, T.; Mekis, A.; Masini, G.; et al. Silicon Photonics R&D and Manufacturing on 300-mm Wafer Platform. *J. Lightwave Technol.* **2016**, *34*, 286–295.
21. Zhao, P.; Zeng, Z.; Yang, N.; Gao, Q.; Tang, B.; Li, Z.; Yan, J.; Jiang, W. Correlation between driving signal reflection on electrodes and performance variation of silicon Mach-Zehnder modulators. *Opt. Express* **2019**, *27*, 35349–35361. [[CrossRef](#)]
22. Yu, H.; Bogaerts, W. An Equivalent Circuit Model of the Traveling Wave Electrode for Carrier-Depletion-Based Silicon Optical Modulators. *J. Lightwave Technol.* **2012**, *30*, 1602–1609. [[CrossRef](#)]
23. Xiong, C.; Gill, D.M.; Proesel, J.E.; Orcutt, J.S.; Haensch, W.; Green, W.M.J. Monolithic 56 Gb/s silicon photonic pulse-amplitude modulation transmitter. *Optica* **2016**, *3*, 1060–1065. [[CrossRef](#)]
24. Anderson, T.W. *An Introduction to Multivariate Statistical Analysis*, 3rd ed.; Wiley-Interscience: Hoboken, NJ, USA; New York, NY, USA, 2003.
25. Saleh, B.E.; Teich, M.C. *Fundamentals of Photonics*; John Wiley & Sons: Hoboken, NJ, USA, 1991.
26. Ghione, G. *Semiconductor Devices for High-Speed Optoelectronics*; Cambridge University: New York, NY, USA, 2009; pp. 388–391.
27. Xu, H.; Li, X.Y.; Xiao, X.; Li, Z.Y.; Yu, Y.D.; Yu, J.Z. Demonstration and Characterization of High-Speed Silicon Depletion-Mode Mach-Zehnder Modulators. *IEEE J. Sel. Top. Quantum Electron.* **2014**, *20*, 23–32. [[CrossRef](#)]
28. Taur, Y.; Ning, T.H. *Fundamentals of Modern VLSI Devices*, 2nd ed.; Cambridge University: New York, NY, USA, 2009; pp. 44–46.
29. Plummer, J.D.; Deal, M.D.; Griffin, P.B. *Silicon VLSI Technology: Fundamentals, Practice and Modeling*; Prentice-Hall: Upper Saddle River, NJ, USA, 2000.
30. Deniel, L.; Weckenmann, E.; Galacho, D.P.; Alonso-Ramos, C.; Boeuf, F.; Vivien, L.; Marris-Morini, D. Frequency-tuning dual-comb spectroscopy using silicon Mach-Zehnder modulators. *Opt. Express* **2020**, *28*, 10888–10898. [[CrossRef](#)]