



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

PH.D. DISSERTATION

A DESIGN OF MULTI-LEVEL SINGLE-ENDED
TRANSMITTERS FOR MEMORY INTERFACES

메모리 인터페이스를 위한 멀티 레벨 단일 종단
송신기 설계

BY

YONG-UN JEONG

AUGUST 2020

DEPARTMENT OF ELECTRICAL ENGINEERING AND
COMPUTER SCIENCE
COLLEGE OF ENGINEERING
SEOUL NATIONAL UNIVERSITY

A DESIGN OF MULTI-LEVEL SINGLE-ENDED TRANSMITTERS FOR
MEMORY INTERFACES

메모리 인터페이스를 위한 멀티 레벨 단일 종단
송신기 설계

지도교수 김 수 환

이 논문을 공학박사 학위논문으로 제출함

2020 년 8 월

서울대학교 대학원

전기컴퓨터 공학부

정 용 운

정용운의 공학박사 학위논문을 인준함

2020 년 8 월

위 원 장 : 정 덕 균 (印)

부위원장 : 김 수 환 (印)

위 원 : 홍 용 택 (印)

위 원 : 최 우 석 (印)

위 원 : 김 진 태 (印)

ABSTRACT

A DESIGN OF MULTI-LEVEL SINGLE-ENDED TRANSMITTERS FOR MEMORY INTERFACES

YONG-UN JEONG

DEPARTMENT OF ELECTRICAL ENGINEERING AND
COMPUTER SCIENCE
COLLEGE OF ENGINEERING
SEOUL NATIONAL UNIVERSITY

Multi-level transmitters for memory interfaces have been presented. The performance gap between processor and memory has been increased by 50% every year, making memory to be a bottle neck of the overall system. To increase memory bandwidth, we have proposed a PAM-4 single-ended transmitter. To compensate for the side effect of the multi-rank memory, we have proposed a reflection-based duobinary transmitter.

The proposed PAM-4 transmitter has the driver, which simultaneously satisfies impedance matching and high linearity. The driver occupies a small area due to a resistorless and inductorless structure. The proposed ZQ calibration for PAM-4 has three calibration points, which allow the transmitter to have accurate impedance and linear output. The ZQ calibration considers impedance variation of both the driver and the receiver. A

prototype has been fabricated in 65nm CMOS process, and the transmitter occupies 0.0333mm^2 . The measured eye has a width of 18.3ps and a height of 42.4mV at 28Gb/s, and the measured energy efficiency is 0.64pJ/b. The measured RLM with the 3-point ZQ calibration is 0.993.

To increase memory density, the stacked die packaging with multiple DRAM die stacked vertically in one package is widely used. However, combined with the center-pad structure, the structure creates stubs that cause short reflections. We have proposed the reflection-based duobinary transmitter to mitigate this problem. The proposed transmitter uses reflection for duobinary signaling. The 2-tap opposite FFE and the slew-rate control are used to increase signal integrity. The measured duobinary eye at 10Gb/s has a width of 63.6ps and a height of 70.8mV while there is no NRZ eye opening. The measured energy efficiency is 1.38pJ/bit.

Keywords: Memory interface, PAM-4 transmitter, ZQ calibration, duobinary transmitter, output driver

Student Number: 2013-20879

CONTENTS

ABSTRACT	1
CONTENTS	3
LIST OF FIGURES	5
LIST OF TABLE	8
CHAPTER 1	1
INTRODUCTION	1
1.1 MOTIVATION.....	1
1.2 THESIS ORGANIZATION.....	8
CHAPTER 2	9
MUTI-LEVEL SIGNALING	9
2.1 PAM-4 SIGNALING	9
2.2 DESIGN CONSIDERATIONS FOR PAM-4 TRANSMITTER	16
2.2.1 LEVEL SEPARATION MISMATCH RATIO (RLM)	17
2.2.2 IMPEDANCE MATCHING	19
2.2.3 PRIOR ARTS	21
2.3 DUOBINARY SIGNALING	24
CHAPTER 3	30
HIGH-LINEARITY AND IMPEDANCE-MATCHED PAM-4 TRANSMITTER	30
3.1 OVERALL ARCHITECTURE	31
3.2 SINGLE-ENDED IMPEDANCE-MATCHED PAM-4 DRIVER	33
3.3 3-POINT ZQ CALIBRATION FOR PAM-4	47
CHAPTER 4	57

REFLECTION-BASED DUOBINARY TRANSMITTER.....57

 4.1 BIDIRECTIONAL DUAL-RANK MEMORY SYSTEM.....58

 4.2 CONCEPT OF REFLECTION-BASED DUOBINARY SIGNALING66

 4.3 REFLECTION-BASED DUOBINARY TRANSMITTER70

 4.3.1 OVERALL ARCHITECTURE70

 4.3.2 EQUALIZATION FOR REFLECTION-BASED DUOBINARY SIGNALING ..72

 4.3.3 2D BINARY-SEGMENTED DRIVER.....75

CHAPTER 5.....77

EXPERIMENTAL RESULTS.....77

 5.1 HIGH-LINEARITY AND IMPEDANCE-MATCHED PAM-4 TRANSMITTER
 77

 5.2 REFLECTION-BASED DUOBINARY TRANSMITTER84

CHAPTER 6.....92

CONCLUSION92

BIBLIOGRAPHY.....94

LIST OF FIGURES

Figure 1.1.1. Various applications of DRAM.	1
Figure 1.1.2. Classification of DRAM according to target performance.	2
Figure 1.1.3. Per-pin bandwidth trends of reported DRAM.	3
Figure 1.1.4. Capacity trends of reported DRAM.	4
Figure 1.1.5. (a) DIMM structure and (b) 3DS memory structure.	5
Figure 2.1.1. Eye diagrams of various PAM signals, (a) PAM-2 (NRZ), (b) PAM-3, and (c) PAM-4.	9
Figure 2.1.2. Comparison of NRZ and PAM-4 signals in (a) time domain and (b) frequency domain.	12
Figure 2.1.3. Block diagram of basic transceiver for PAM-4 interface.	14
Figure 2.2.1.1. Definition of level separation mismatch ratio, RLM.	17
Figure 2.2.1.1. Impedance discontinuity points causing reflection in signal interface.	19
Figure 2.2.3.1. Current-mode differential driver for PAM-4 signaling.	21
Figure 2.2.3.2. Voltage-mode driver with series resistors for PAM-4 signaling.	22
Figure 2.3.1. Example of NRZ data and Duobinary data.	24
Figure 2.3.2. Comparison of NRZ data and Duobinary data in frequency domain.	25
Figure 2.3.3. Block diagram of basic transceiver for duobinary signaling.	26
Figure 2.3.4. Example of NRZ data and Duobinary data with frequency-dependent channel loss.	27
Figure 2.3.5. Eye diagrams of (a) NRZ and (b) duobinary signaling.	28
Figure 3.1.1. Overall block diagram of proposed PAM-4 transmitter.	31
Figure 3.2.1. N-over-N driver for NRZ signaling.	33
Figure 3.2.2. Voltage-mode PAM-4 driver using N-over-N structure.	34
Figure 3.2.3. Operation of generating four levels for PAM-4 driver using N-over-N structure, assuming ideal MOSFETs are used.	36

Figure 3.2.4. Impedance of N-over-N driver according to output voltage.	37
Figure 3.2.5. (a) Operation of PAM-4 driver using N-over-N structure, and (b) its output eye diagram with impedance variation.	39
Figure 3.2.6. Proposed impedance-matched PAM-4 driver.	40
Figure 3.2.7. Operation of proposed IM-PAM-4 driver.	41
Figure 3.2.8. Circuit implementation of proposed IM-PAM-4 driver.	44
Figure 3.2.9. (a) Block diagram and (b) encoding logic of encoder for IM-PAM-4 driver.	45
Figure 3.3.1. Conventional ZQ calibrations of (a) pull-up driver and (b) pull-down driver for NRZ signaling.	47
Figure 3.3.2. Overall architecture of proposed 3-point ZQ calibration.	49
Figure 3.3.3. First loop operation of proposed ZQ calibration at $1/6 \cdot V_{DDQ}$	50
Figure 3.3.4. Second loop operation of proposed ZQ calibration at $1/6 \cdot V_{DDQ}$	52
Figure 3.3.5. First loop operation of proposed ZQ calibration at $1/3 \cdot V_{DDQ}$	53
Figure 3.3.6. Second loop operation of proposed ZQ calibration at $1/3 \cdot V_{DDQ}$	54
Figure 3.3.7. First loop operation of proposed ZQ calibration at $1/2 \cdot V_{DDQ}$	55
Figure 4.1.1. (a) Stacked-die packaging and (b) center-pad structure for memory.	58
Figure 4.1.2. Overall architecture of stacked die packaging with center-pad structure.	59
Figure 4.1.3. Simulated eye diagram with 2-rank memory system.	60
Figure 4.1.4. (a) READ and (b) WRITE operation of 2-rank memory system.	61
Figure 4.1.5. (a) First, (b) second and (c) third arrived signal.	62
Figure 4.1.6. Lattice diagram for 2-rank memory system.	64
Figure 4.2.1. Simulated eye diagrams at data-rates with 1-UI near $2 \cdot T_{RDL}$ in 2-rank memory system.	67
Figure 4.2.2. Transition areas of eye diagrams at data-rates with 1-UI near $2 \cdot T_{RDL}$ in 2-rank memory system.	68
Figure 4.3.1.1. Overall block diagram of proposed reflection-based duobinary transmitter.	71
Figure 4.3.1.1. Simulated single-bit response with and without equalization at $1\text{-UI} >$	

$2 \cdot T_{RDL}$	72
Figure 4.3.2.2. Simulated single-bit response with and without equalization at (a) $1-UI \approx 2 \cdot T_{RDL}$ and (b) $1-UI < 2 \cdot T_{RDL}$	74
Figure 4.3.3.1. Architecture of 2D binary-segmented driver	75
Figure 4.3.3.2. The required number of segmentations to independently control two characteristics	76
Figure 5.1.1. Chip micrograph of the proposed PAM-4 transmitter	77
Figure 5.1.2. Measurement environment for the proposed PAM-4 transmitter	78
Figure 5.1.3. Measured RLM with (a) one-point calibration and (b) three-point calibration.	79
Figure 5.1.4. Measured eye diagram of the proposed PAM-4 transmitter at 28Gb/s with PRBS-7 pattern	80
Figure 5.1.5. Bathtub curves for the proposed PAM-4 transmitter at 28Gb/s	81
Figure 5.1.6. Power breakdown of the proposed PAM-4 transmitter at 28Gb/s	82
Figure 5.1.1. Chip micrograph of the proposed reflection-based duobinary transmitter ..	84
Figure 5.2.2. Measurement environment for the proposed reflection-based duobinary transmitter	85
Figure 5.2.3. Measured eye diagram at 8Gb/s (a) without equalization and (b) with equalization	87
Figure 5.2.4. Measured eye diagram at 9Gb/s (a) without equalization and (b) with equalization	88
Figure 5.2.5. Measured eye diagram at 10Gb/s (a) without equalization and (b) with equalization	89
Figure 5.2.6. Bathtub curves for the proposed reflection-based duobinary transmitter at 10Gb/s	90

LIST OF TABLE

Table 5.1.1 Performance Summary and Comparison with Other Multi-Level Transmitters
..... 83

Table 5.2.1 Performance Summary and Comparison with Transmitters for Multi-Drop
Systems
..... 91

CHAPTER 1

INTRODUCTION

1.1 MOTIVATION

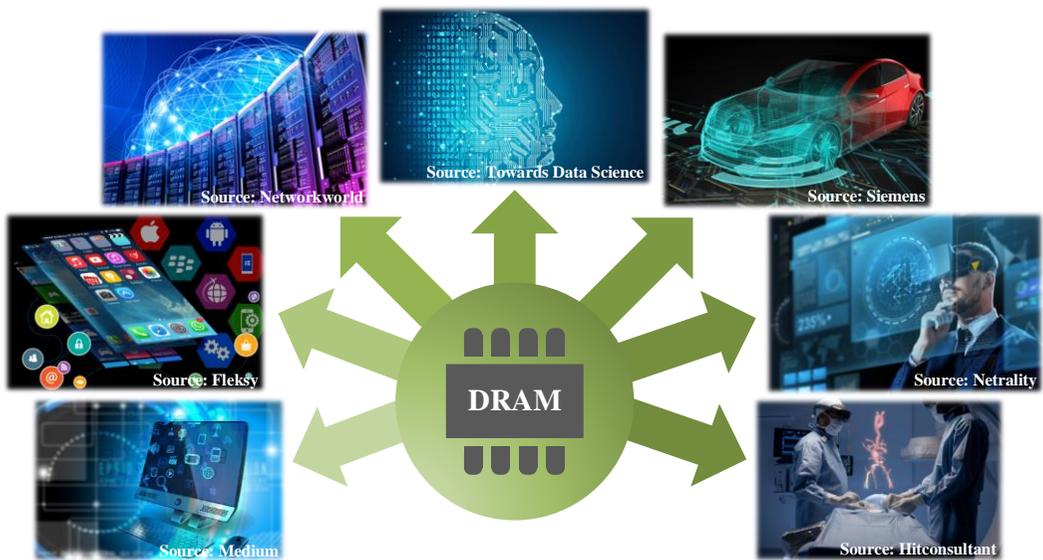


Figure 1.1.1. Various applications of DRAM.

Dynamic random-access memory (DRAM) is a typical storage device that stores digital data in the form of charge in each memory cell. The rapid development of big-data-based technologies and numerous applications such as mobile devices, internet of things

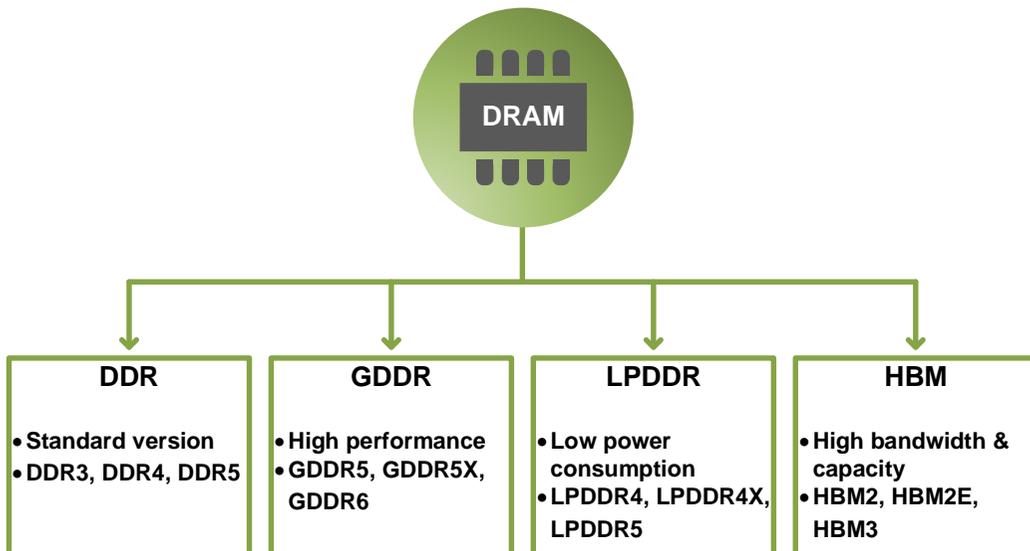


Figure 1.1.2. Classification of DRAM according to target performance.

(IoT), autonomous vehicles, consumer electronics, personal computers and gaming consoles requires higher bandwidth and higher capacity DRAM, as shown in Figure 1.1.1. In particular, as the performance difference between processor and DRAM increases over time, DRAM becomes the bottle neck of the overall system.

DRAM has been developing various types of DRAM suitable for each application in order to satisfy various demands. As shown in Figure 1.1.2, DRAMs are largely categorized as general purpose double data rate (DDR) synchronous dynamic random-access memory (SDRAM), graphic DDR (GDDR) SDRAM for fast transfer rates, and low-power DDR (LPDDR) SDRAM for low power consumption, and high-bandwidth memory (HBM) for

Memory Per-Pin Bandwidth Trends

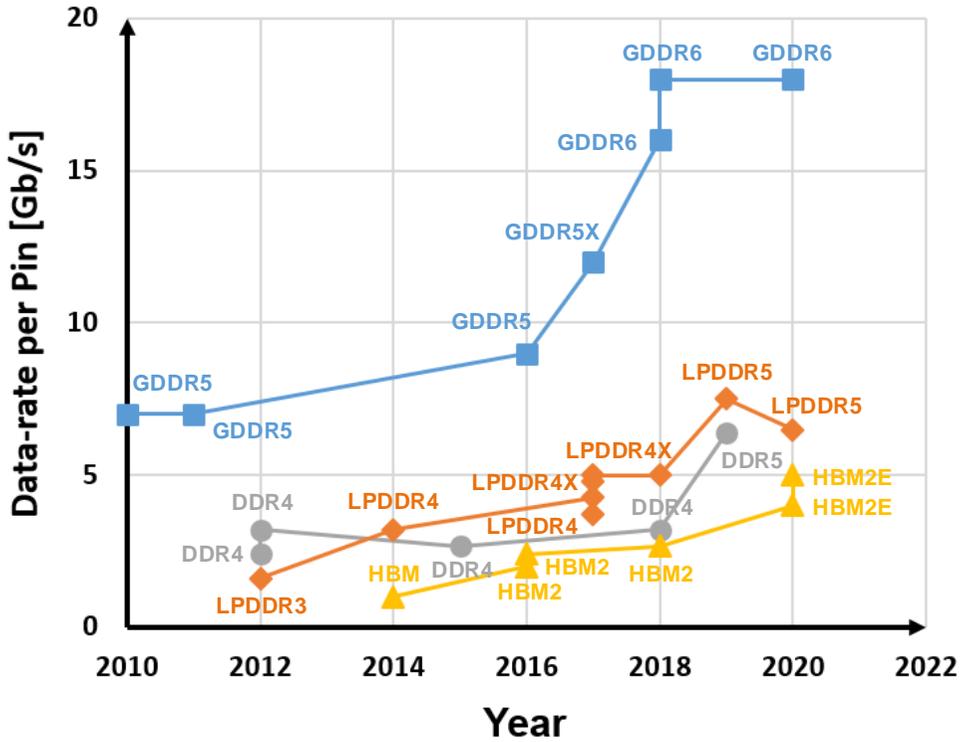


Figure 1.1.3. Per-pin bandwidth trends of reported DRAM.

high bandwidth and capacity. They satisfy the performance required by each application, for example, low power required by mobile devices, high transmission speed required by high performance graphics, and high bandwidth and capacity required by servers.

Figure 1.1.3 and Figure 1.1.4 show trends of bandwidth and capacitance of various types of DRAM presented at conferences since 2010. GDDR has higher bandwidth per pin

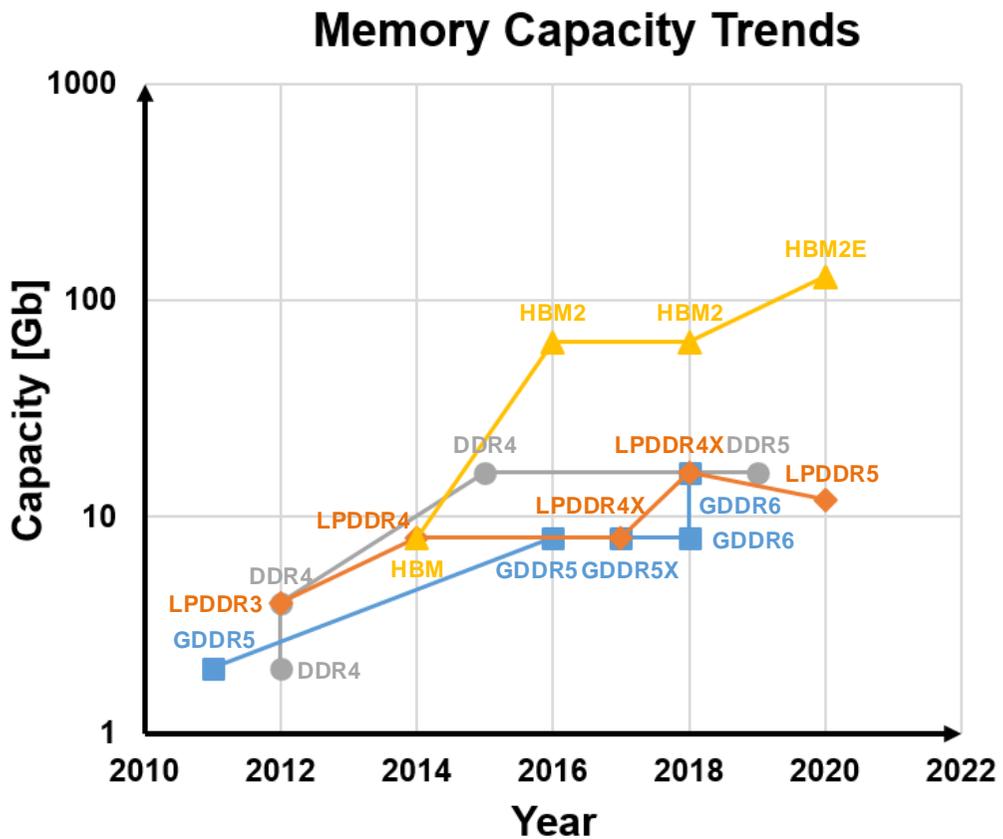
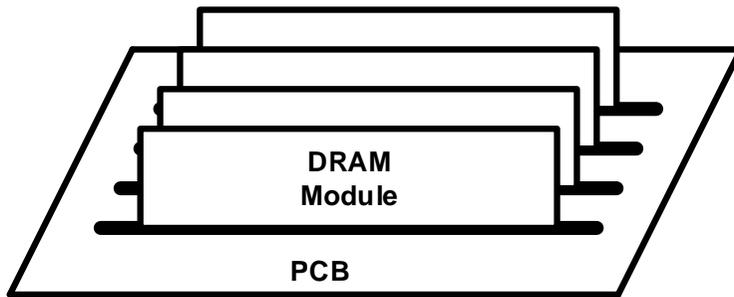
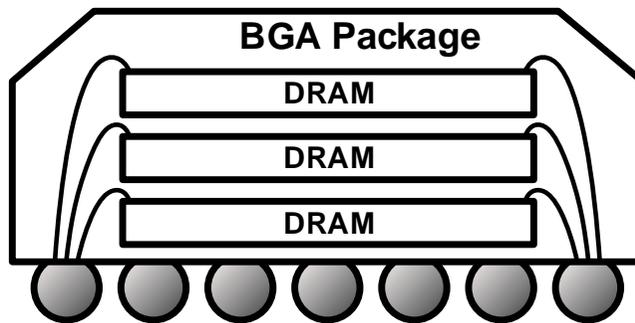


Figure 1.1.4. Capacity trends of reported DRAM.

than other types of DRAM by focusing on data rate per in/out (I/O) pin rather than density [1.1.1]. HBM has a structure that greatly increases the number of I/O in order to increase the total bandwidth, and it has the highest total bandwidth and memory capacity although the performance per I/O is low [1.1.2]. Each DRAM is optimized for the target performance, but all applications need to improve all these performances because the direction of development is toward higher bandwidth, higher capacity, and lower power consumption.



(a)



(b)

Figure 1.1.5. (a) DIMM structure and (b) 3DS memory structure.

DRAM performance has been increased through three major methods. The first is the development of the process. The process scaling and the developed performance of the memory cell achieve an increase in the density and the bandwidth of DRAM. This direction has the advantage of obtaining a fundamental performance increase of DRAM compared to other methods, and thus, continuous progress is being made every year. However, there is a limit to the speed of performance development through this. This is because just scaling the process twice will not double the DRAM density and I/O performance. The second way

to improve performance is parallelization. As shown in Figure 1.1.5, typical methods of parallelization are dual in-line memory module (DIMM) with multiple DRAM modules mounted on one circuit board, and three-dimensional structure (3DS) memory with multiple DRAM die stacked vertically in one package [1.1.3]. Parallelization is a solution that allows multiple DRAMs to be connected in parallel to obtain higher density as a way to overcome the limitations of density increase from process development. However, this method also has limitations such as degraded signal integrity from the multi-drop structure and technical limitations of die stacking [1.1.4]. The third method for DRAM development is to increase the circuit performance. This is a method to increase the performance through the development of the circuit structure of DRAM, and overcome the above-mentioned problems arising from the process development and parallelization with the development of circuit design. In particular, as the data-rate per pin that determines I/O performance increases, the importance of impedance matching with channels and equalization technology to overcome frequency-dependent loss is increasing, and these are typical issues to be solved in DRAM circuit design [1.1.5]. In addition, the circuit development overcomes the signal integrity deterioration caused by the multi-drop structure due to parallelization, obtaining a high density of DRAM. All three DRAM development directions mentioned above have advantages, but there are also limitations. Therefore, all three methods must be performed simultaneously, and higher performance can be achieved by solving each limitation in another method.

This thesis aims to increase circuit performance, one of the ways to improve DRAM performance. Two types of multi-level signaling for DRAM have been proposed as

solutions to satisfy the explosively increasing demand for DRAM and the increasing demand for performance. First, a PAM-4 transmitter is proposed to increase the data rate per pin. The proposed PAM-4 driver has a resistorless and inductorless structure, which is suitable for memory, occupying a small area and achieving impedance matching and high linearity. A ZQ calibration scheme for PAM-4 signaling has also been proposed. This calibration allows the proposed PAM-4 driver to achieve impedance matching and high linearity automatically. Second, a reflection-based duobinary transmitter for bidirectional multi-rank memory was proposed. The proposed transmitter uses reflection generated by the multi-rank memory structure for duobinary signaling. The proposed transmitters are verified through analysis and measurement.

1.2 THESIS ORGANIZATION

This thesis is organized as follows: in Chapter 2, two multi-level signaling, PAM-4 and duobinary are introduced; in Chapter 3, the thesis presents the PAM-4 single-ended transmitter with high-linearity and impedance matched driver and 3-point ZQ calibration; in Chapter 4, a reflection-based duobinary transmitter is presented; in Chapter 5, experimental results are presented; and in Chapter 6, the thesis is summarized with the discussion of contribution.

CHAPTER 2

MULTI-LEVEL SIGNALING

2.1 PAM-4 SIGNALING

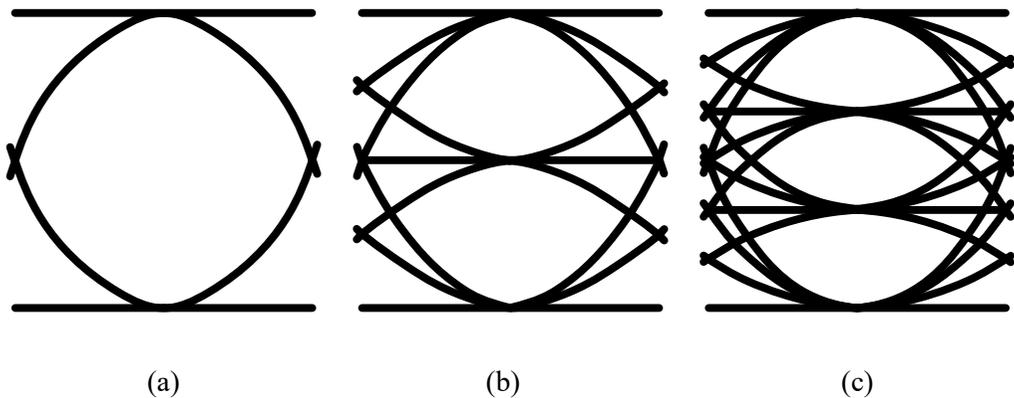


Figure 2.1.1. Eye diagrams of various PAM signals, (a) PAM-2 (NRZ), (b) PAM-3, and (c) PAM-4.

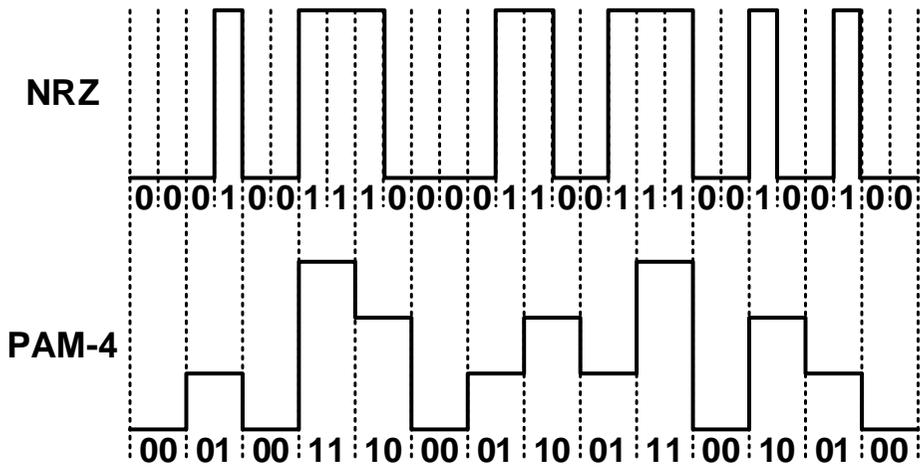
As the demand for high bandwidth increases, non-return-to-zero (NRZ), which transmits 1-bit information of 0 or 1 in a single symbol, has a limit for bandwidth improvement. As one of the solutions, pulse-amplitude modulation (PAM) contains more

than 1-bit information in one symbol through amplitude modulation of the transmitted signal [2.1.1]. NRZ signaling is also a type of PAM, PAM-2 that contains two information, 0 and 1, in amplitude. Figure 2.1.1 shows eye diagrams of various PAM signals. Figure 2.1.1 (a) shows the eye diagram of PAM-2, NRZ, which contains 1-bit information using 2 signal levels in one symbol. Figure 2.1.1 (b) shows the eye diagram of PAM-3 containing three levels, or 1.5-bits, in one symbol. PAM-3 can transmit 1.5 times more information than NRZ within the same time, and when the maximum signal swing is the same, it has an interval between signal levels about two times lower than NRZ. This means that the eye height is also reduced by about two times. Figure 2.1.1 (c) shows the eye diagram of the PAM-4 signal. Since PAM-4 has four signal levels in one symbol, 2-bit information can be simultaneously transmitted.

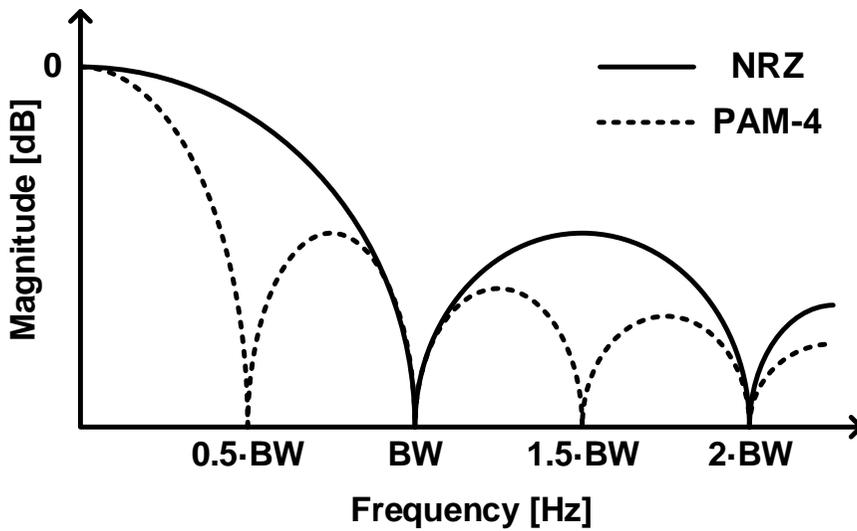
In the case of PAM-3 encoding, the digital data of 1-bit / symbol inside the chip must be encoded as 1.5-bit / symbol, so the relationship between the input and output data of the encoder is not an integer. This leads to the complexity of the encoder structure or loss of some of the information to simplify the structure [2.1.2]. On the other hand, PAM-4 encoding is very simple compared to PAM-3 because the ratio of information between input and output is an integer when converting 1-bit/symbol data to 2-bit/symbol data. PAM-8, which has 8 levels in one symbol, also has 3-bit/symbol data, so the encoder structure is relatively simple. However, since the interval between signal levels is 7 times smaller compared to NRZ, it is inevitably vulnerable to noise.

PAM-4 has an interval between signal levels 3 times smaller than NRZ when the same maximum signal swing is assumed, as shown in Figure 2.1.1 (c). This means that the PAM-

4 signal has a signal-to-noise-ratio (SNR) of about 9.5dB lower than the NRZ signal [2.1.3]. Moreover, PAM-4 not only reduces the signal level, but also causes various signal transitions due to the increased number of signal levels. This means that each eye of PAM-4 receives three times more inter-symbol interference (ISI) [2.1.4]. Therefore, the PAM-4 signal has reduced vertical margins of eyes compared to NRZ. Lower vertical margins make signals more susceptible to process, supply voltage, and temperature (PVT) variations, crosstalk, reflection, and other noises. Therefore, careful consideration for small eye height is required for PAM-4 design.



(a)



(b)

Figure 2.1.2. Comparison of NRZ and PAM-4 signals in (a) time domain and (b) frequency domain.

Figure 2.1.2 shows the waveforms in the time domain and power spectrum densities in the frequency domain of NRZ and PAM-4 signals. Assuming the same data-rate, PAM-4 has twice the unit interval (UI) length compared to NRZ. This means that when comparing two signals in the frequency domain, the bandwidth of the PAM-4 signal is reduced by half compared to the NRZ signal. The transmission line, also called a channel constituting the interface, generally has frequency-dependent characteristics. Channels have low loss for low frequencies and high loss for high frequencies. Therefore, a PAM-4 signal having lower bandwidth at the same data-rate has an advantage in a frequency-dependent channel [2.1.5]. In particular, when the difference between the channel loss at the PAM-4 bandwidth and the channel loss at the NRZ bandwidth is 9.5dB or more, it is advantageous to use PAM-4 signaling because the SNR of 9.5dB is compensated from the channel.

From the perspective of the circuit in the chip, the advantages of PAM-4 signaling are more evident. Figure 2.1.3 shows the basic transceiver for the PAM-4 interface. The PAM-4 transmitter consists of a serializer that serializes multiple low-speed parallel data into high-speed serial data, an encoder that generates 2-bit data, MSB, and LSB for PAM-4 signaling, and a PAM-4 driver. The simplest encoder to create the 2-bit data of PAM-4 is to remove the last 2:1 serializer from the serialzier for NRZ [2.1.6]. For example, if an N:1 serializer is used for NRZ, an N:2 serializer is used for PAM-4. The PAM-4 receiver consists of an amplifier to compensate for the reduced signal due to the channel, a sampler to restore digital data, a decoder to decode PAM-4 data into NRZ data, and a deserializer to make high-speed serial data into multiple low-speed parallel data. The simplest decoder,

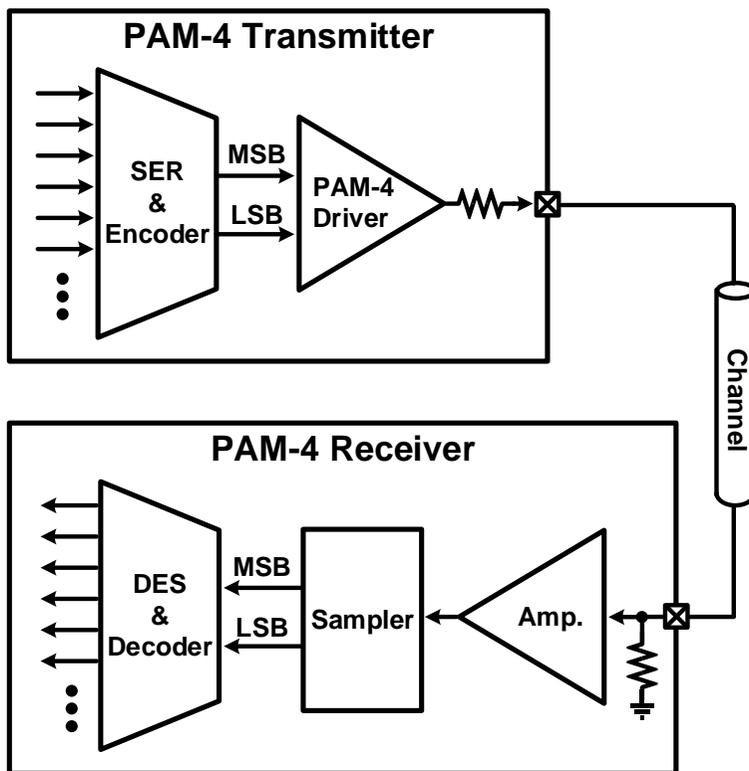


Figure 2.1.3. Block diagram of basic transceiver for PAM-4 interface.

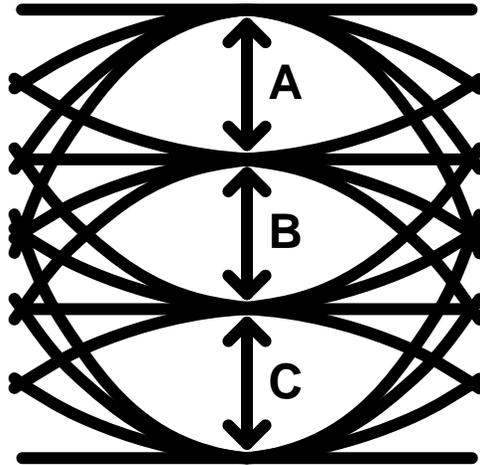
similar to the PAM-4 encoder, is to remove the first 1:2 deserialzier from the deserialzier [2.1.6]. The main advantage of PAM-4 in terms of internal circuitry is its operating speed. Assuming the same data-rate, PAM-4 has a 1-UI length twice that of NRZ, so the operating speed of the internal circuit is half that of NRZ. In other words, when using circuits with the same operating speed, PAM-4 can interface with twice the data-rate compared to NRZ.

In particular, in the case of memory, the process for memory is difficult to increase the operating speed compared to other processes. Therefore, PAM-4 is an attractive solution for high bandwidth in the memory interface.

2.2 DESIGN CONSIDERATIONS FOR PAM-4 TRANSMITTER

When designing a PAM-4 transmitter, the biggest challenge compared to the NRZ transmitter is the smaller vertical margin. As mentioned before, PAM-4 signal has 3 times smaller difference between signal levels and 3 times larger ISI compared to NRZ, leading to more than 3 times smaller eyes. This means that the receiver must have at least 3 times higher sensitivity to recognize the eyes. Also, small eyes make the signal more susceptible to noise, crosstalk and PVT variations. In order to reduce these burdens, it is important to secure the maximum vertical margin of the eye. PAM-4 has 3 eyes in one symbol, and determining the performance of PAM-4 is the smallest eye. This means that the performance of PAM-4 is maximized when three eyes have the same height. To have a uniform eye height, it is important to have four accurate signal levels.

2.2.1 LEVEL SEPARATION MISMATCH RATIO (RLM)



$$\text{RLM} = \frac{\min(A,B,C)}{(A+B+C)/3}$$

Figure 2.2.1.1. Definition of level separation mismatch ratio, RLM.

Figure 2.1.1.1 shows the definition of level separation mismatch ratio (RLM), an indicator of the linearity of PAM-4 [2.2.1.1]. RLM is the ratio of the smallest of the intervals between levels and the average of the intervals. RLM is in the range greater than 0 and smaller than 1, and the closer the RLM is to 1, the more the signal levels are uniform, meaning that PAM-4 is linear. On the other hand, as the RLM approaches 0, it means that the smaller interval is smaller than the average, and that the PAM-4 is non-linear. The rest

of the gaps are wider, but in the end, the eye in the smallest interval is dominant because it produces the most errors. Therefore, the closer the RLM is to 0, the lower the performance.

There are many factors that degrade RLM. Above all, the driver is the most important because it makes 4 levels of PAM-4 and is the final stage of the transmitter. Therefore, mismatch of each component constituting the driver makes the RLM low. Other factors such as lower supply voltage due to drop and fluctuation of the output voltage in the pre-driver can also cause the RLM to be lowered.

2.2.2 IMPEDANCE MATCHING

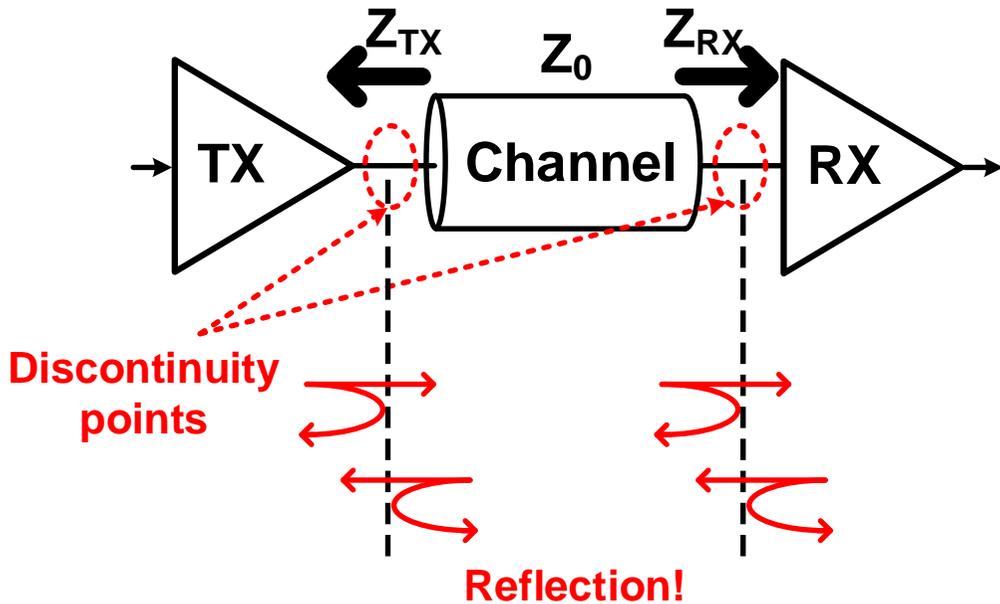


Figure 2.2.1.1. Impedance discontinuity points causing reflection in signal interface.

Figure 2.2.2.1 shows the importance of impedance matching. Many chip-to-chip communications transmit signals over transmission lines, also called channels. The transmission line has a characteristic impedance, and reflection occurs at the point where the impedance changes. The parts that make these discontinuities are typically the point where the transmitter output meets the channel, the point where the channel meets the receiver input, stubs that can occur on a channel trace such as a PCB, the point where the channel meets another channel, and connectors such as SMA on the signal trace. Reflection at these discontinuities is undesired noise and degrades signal integrity. The impedance of

the transmitter and receiver is the impedance seen from the transmitter output and the receiver input, respectively. In the case of a receiver, the input through the channel generally enters the gate of the MOSFET of the receiver's first circuit, so to obtain the desired impedance, the impedance is matched using a resistor in parallel with the gate of the MOSFET. Since the impedance value of the gate of the MOSFET is very large, the impedance is determined only by the resistor connected in parallel, which makes the impedance matching with the channel in the receiver relatively easy. On the other hand, in the case of a transmitter, the impedance of the MOSFET affects the transmitter output impedance because the output node is connected to the sources or drains of the MOSFETs that constitute the last driver [2.2.2.1]. In the case of the current-mode driver, which is a type of driver, the total impedance is very large because this MOSFET is connected in series with the current source. Therefore, like receiver, impedance matching with the channel can be facilitated by using parallel resistors. However, current-mode drivers are difficult to configure as single-ended signaling and use differential signaling [2.2.2.2]. Differential signaling uses twice the number of output pins compared to single-ended signaling, so the cost of additional pins is high when many output drivers are used such as memory. Therefore, a voltage-mode driver with high pin efficiency is used in memory rather than a current-mode driver. In the case of voltage-mode drivers, MOSFETs connected to the output node are usually connected to supply voltage, ground, or regulated voltage [2.2.2.3]. This means that the impedance of the MOSFET affects impedance matching. Therefore, many voltage-mode drivers try to set the impedance of the MOSFETs to the desired value for impedance matching [2.2.2.3].

2.2.3 PRIOR ARTS

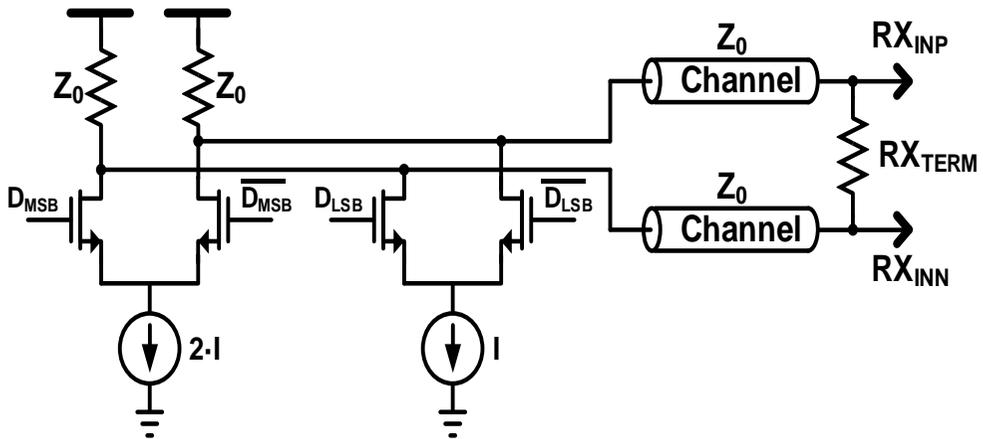


Figure 2.2.3.1. Current-mode differential driver for PAM-4 signaling.

The easiest way to satisfy the high linearity of pam-4 levels and impedance matching is to use a current-mode differential driver, as shown in Figure 2.2.3.1 [2.2.2.2]. The current-mode driver has stable output because it uses stable current sources. Therefore, high linearity of levels and impedance matching can be easily satisfied. However, as mentioned earlier, differential signaling has a high cost using many pins compared to single-ended signaling. In addition, the current-mode driver theoretically consumes 2 to 4 times higher power than the voltage-mode driver. The high cost and low energy efficiency make the current-mode driver unsuitable for memory applications. Moreover, for a stable current source, the length of the MOSFET constituting the current source must be designed

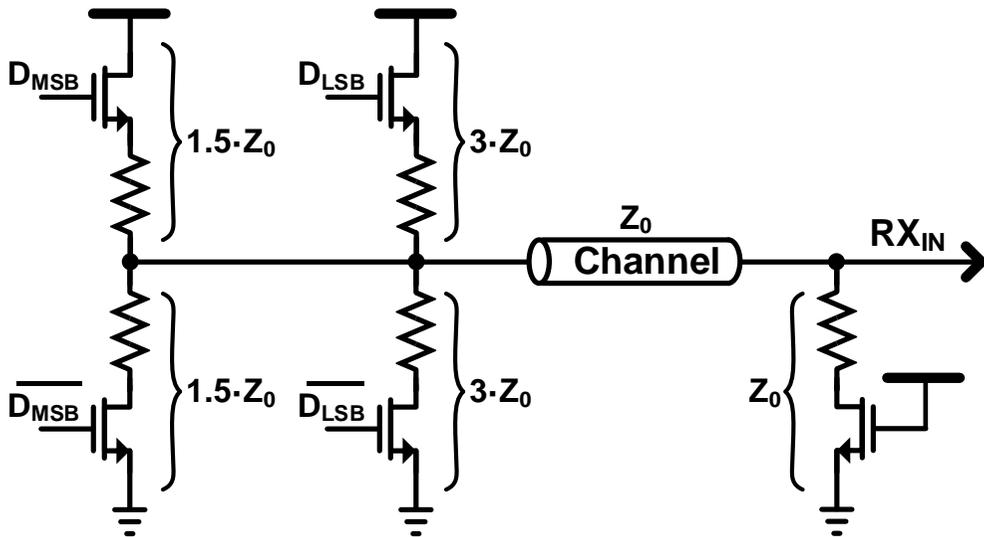


Figure 2.2.3.2. Voltage-mode driver with series resistors for PAM-4 signaling.

large. Therefore, a current-mode driver with stable current sources occupies a large area, which makes it unsuitable for memory applications where density is important.

On the other hand, as shown in Figure 2.2.3.2, the voltage-mode driver has less power consumption than the current-mode driver, and the pin efficiency is high because a single-ended structure can be easily obtained. However, since impedances of the MOSFETs constituting the driver affect the impedance of the transmitter output, it is necessary to accurately maintain the impedance of the MOSFET for impedance matching. A common method to reduce the impedance variation of MOSFETs is to use series passive resistors [2.2.3.1]. These allow the transmitter output impedance to be determined by the sum of the

impedance of the MOSFET and the resistance of the resistor. Therefore, the larger the proportion of the passive resistance, which has a relatively small impedance variation, in the total impedance, the smaller the variation in the transmitter output impedance. However, this also means that the impedance value that the MOSFET has is small. Small impedance MOSFETs occupy a large area. Moreover, although the series resistors reduced the total impedance variation, the impedance variation of the MOSFET still remains, and the resistors also have resistance variation due to PVT variations. Therefore, this structure improves impedance matching and linearity with series resistors at the expense of area, but still remains.

2.3 DUOBINARY SIGNALING

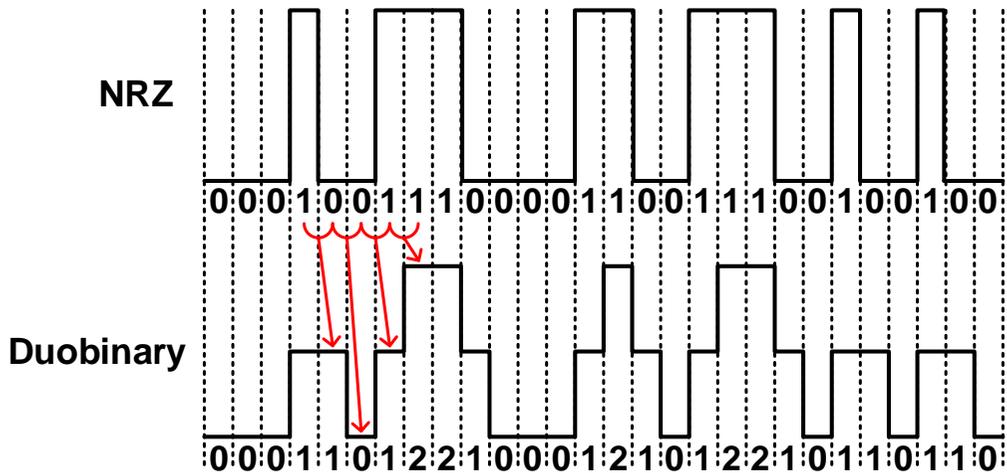


Figure 2.3.1. Example of NRZ data and Duobinary data.

Duobinary signaling uses encoding that combines current data and previous data [2.3.1], which can be expressed as follows:

$$r[n] = x[n] + x[n - 1], \quad (2.3.1)$$

where $r[n]$ represents n th encoded data and $x[n]$ represents n th original data. The original data, $x[n]$ is NRZ data consisting of 0 and 1, and encoded data, $r[n]$ is 3-level data consisting of 0, 1, and 2. Figure 2.3.1 shows an example of duobinary encoding. The advantage of encoding from 2 levels of NRZ data to 3 levels of duobinary data is the bandwidth of the

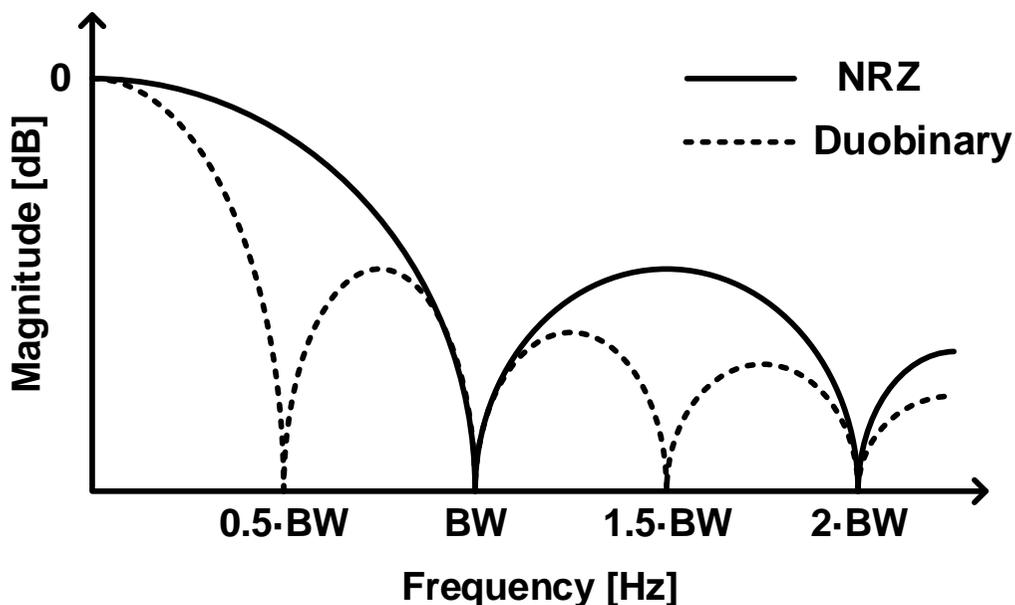


Figure 2.3.2. Comparison of NRZ data and Duobinary data in frequency domain.

data. Figure 2.3.2 shows the comparison of NRZ data and duobinary data in frequency domain. Compared to the bandwidth of NRZ data, the duobinary bandwidth has half the bandwidth. Low bandwidth makes duobinary data insensitive to channels with frequency-dependent loss compared to NRZ data. This also means that duobinary signaling is suitable for high loss channel environments.

The biggest feature of the duobinary is not in the encoding itself, but in the encoding method, which uses the frequency-dependent characteristic of the channel [2.3.2]. Figure 2.3.3 shows a block diagram of a basic transceiver using duobinary signaling. Other

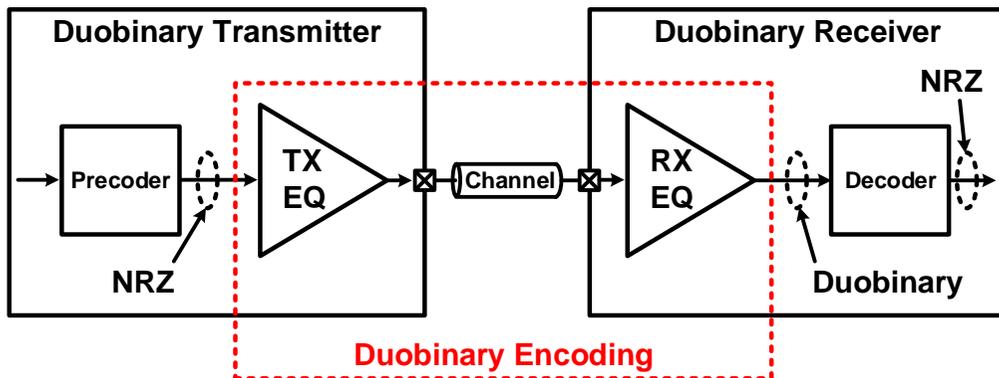


Figure 2.3.3. Block diagram of basic transceiver for duobinary signaling.

signals such as NRZ signaling and PAM-4 signaling try to amplify the high frequency components of the signal to compensate for channel loss. On the other hand, duobinary signaling includes the frequency-dependent loss characteristic of the channel in the function of data encoding. The frequency-dependent loss of the channel attenuates the high-frequency components of the data, especially slowing the transition of the data. Slow data transitions cause data to go beyond 1-UI and affect other data areas, which is a type of ISI. Duobinary uses this ISI where the previous data affects the next data due to channel loss. The equalizers of the transmitter and the receiver help this ISI to be the desired amount for the duobinary encoding. Figure 2.3.4 shows duobinary data including the frequency-dependent loss characteristics of the channel. Three levels of duobinary data are created using the slow transition due to channel loss. Figure 2.3.5 shows a comparison of the NRZ

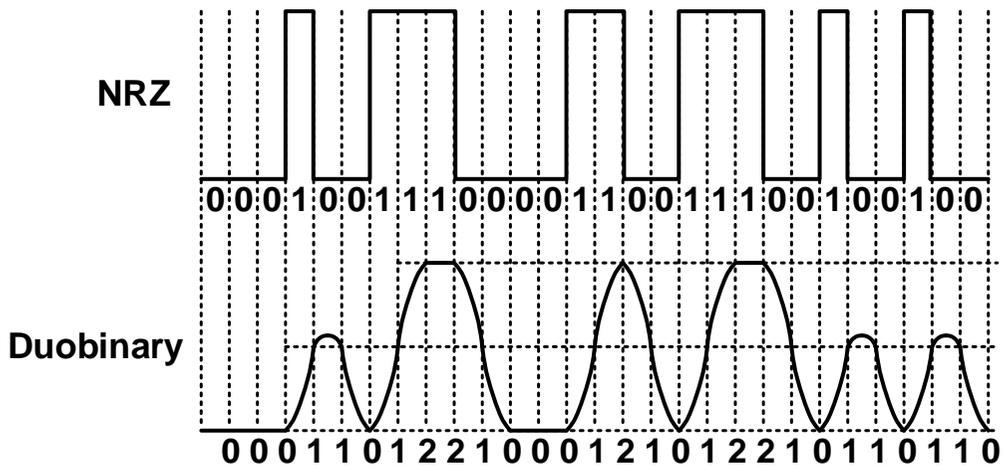


Figure 2.3.4. Example of NRZ data and Duobinary data with frequency-dependent channel loss.

and duobinary eye diagrams.

From Eq. (2.3.1), decoding logic to convert duobinary data to NRZ data can be obtained, which can be expressed as follows:

$$x[n] = r[n] - x[n - 1]. \quad (2.3.2)$$

This expression indicates that the previous data, $x[n-1]$, is used to know the current data, $x[n]$. However, if an error occurs in the data, the error data causes the next data to be judged incorrectly, and these errors continue to accumulate. To prevent cumulative errors,

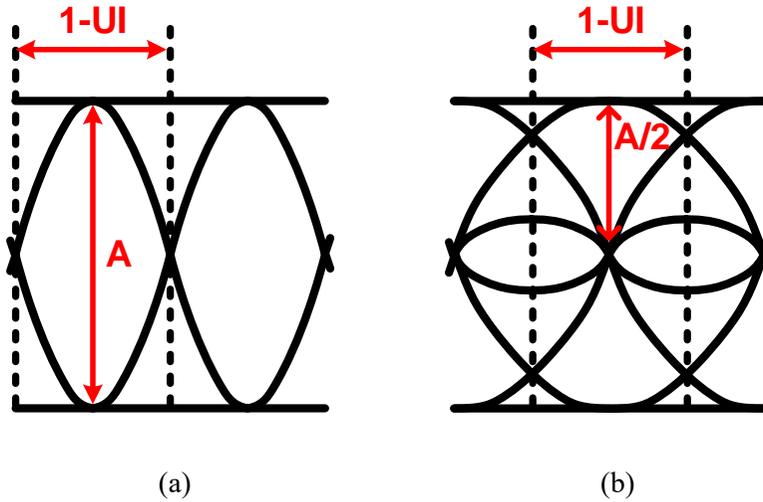


Figure 2.3.5. Eye diagrams of (a) NRZ and (b) duobinary signaling.

a precoder is used in duobinary signaling [2.3.3]. The precoder is located in the transmitter before the data is encoded to duobinary, and its behavior can be expressed as follows:

$$y[n] = y[n - 1] \oplus x[n]. \quad (2.3.3)$$

where $y[n]$ denotes the encoded data by the precoder and \oplus denotes XOR function. The duobinary signal when using the precoder can be expressed as follows:

$$r[n] = y[n] + y[n - 1] = (y[n - 1] \oplus x[n]) + y[n - 1]. \quad (2.3.4)$$

This results also can be interpreted as follows:

$$x[n] = \begin{cases} 1 & \text{if } r[n] = 1 \\ 0 & \text{if } r[n] = 0 \text{ or } 2 \end{cases} \quad (2.3.5)$$

This means that the previous data is not needed to decide the current data.

CHAPTER 3

HIGH-LINEARITY AND IMPEDANCE-MATCHED PAM-4 TRANSMITTER

To alleviate the issues about PAM-4 transmitter discussed in Chapter 2.1 and 2.2, we have presented a single-ended PAM-4 transmitter with a high-linearity and impedance-matched driver and 3-point ZQ calibration. The proposed PAM-4 driver satisfies both the high-linearity and impedance matching simultaneously by an additional pull-up driver. This does not use any resistors or inductors, leading to a small area. The ZQ calibration for PAM-4 operates at three calibration points to compensate for the impedance variation according to the output levels of PAM-4. This calibration takes impedance information from the external reference resistor and the DQ driver to perform accurate impedance calibration.

3.1 OVERALL ARCHITECTURE

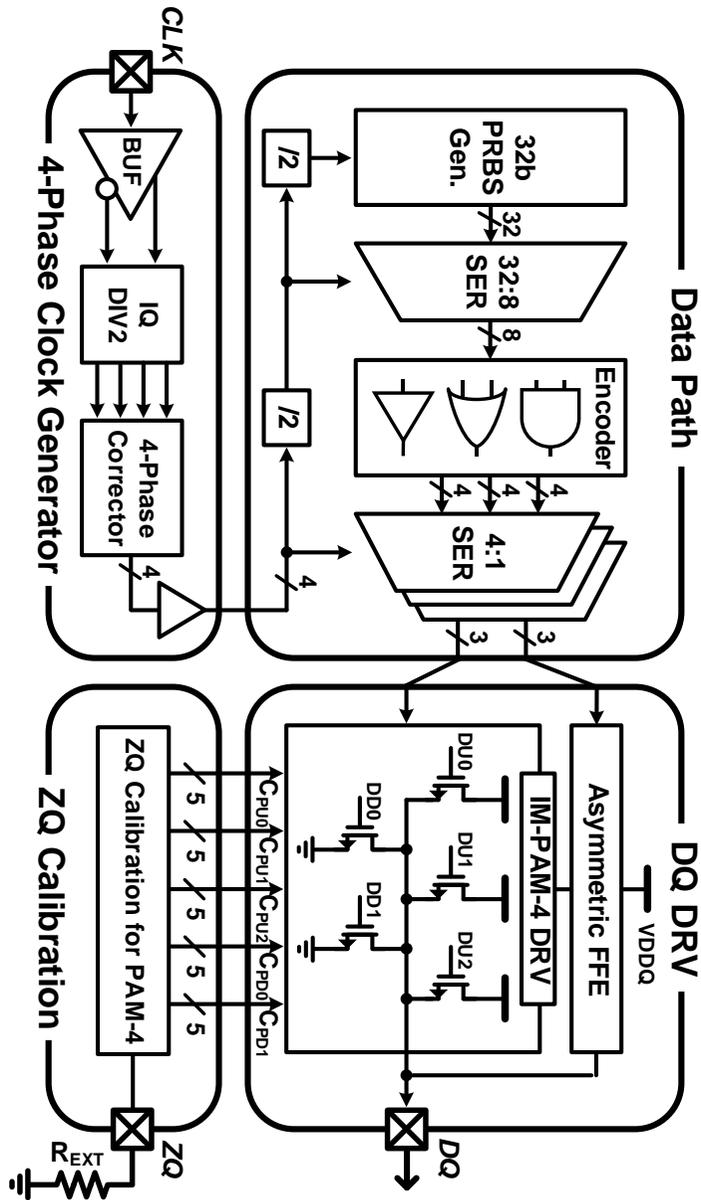


Figure 3.1.1. Overall block diagram of proposed PAM-4 transmitter.

Figure 3.1.1 shows the overall block diagram of the proposed single-ended PAM-4 transmitter. The transmitter has a quarter-rate structure using quadrature clocks, which allows the transmitter to have a lower clock frequency. A 32-bit pseudo random binary sequence (PRBS) generator is used for the test. The generated 32-bit PRBS pattern is serialized from low-speed parallel data to high-speed series data through a 32:8 serializer. The serialized data is delivered to PAM-4 driver through an encoder and a 4:1 serializer. By placing the encoder in front of the 4:1 serializer, the jitter generated by the encoder is eliminated from the 4:1 serializer. The proposed impedance-matched PAM-4 (IM-PAM-4) driver satisfies both impedance matching and high linearity to obtain high PAM-4 signal integrity. The ZQ calibration for PAM-4 performs impedance calibration so that the driver has the correct impedance.

3.2 SINGLE-ENDED IMPEDANCE-MATCHED PAM-4 DRIVER

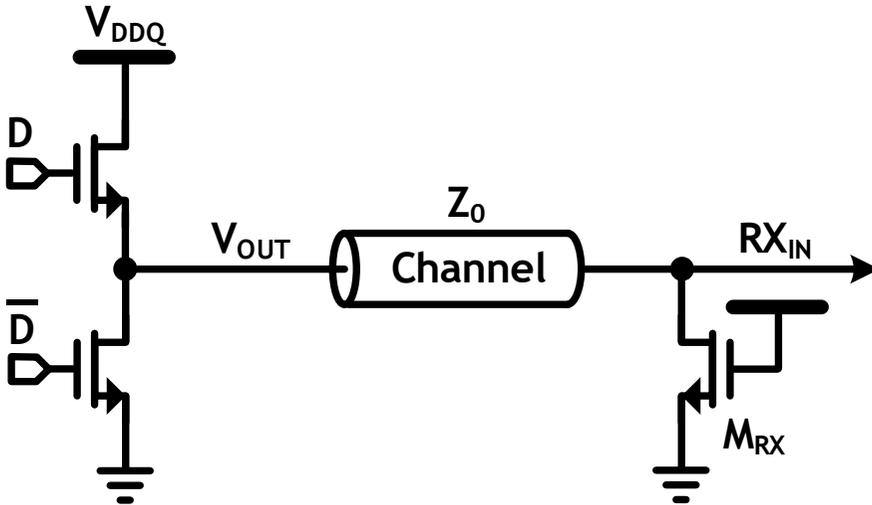


Figure 3.2.1. N-over-N driver for NRZ signaling.

Figure 3.2.1 shows the N-over-N driver for NRZ signaling used in memory. This driver consists of one pull-up driver and one pull-down driver consisting of NMOS only. There are two main advantages of using NMOS rather than PMOS as a pull-up driver. First, using NMOS instead of PMOS, the same impedance can be achieved with a smaller size. Second, the small size reduces the I/O capacitive load (C_{IO}) of the output node, which allows the transmitter to achieve higher bandwidth. Pull-up NMOS makes the driver difficult to operate at high supply voltage, but rather, the advantages of this structure at low voltage make the driver more suitable for low-power memory applications.

Figure 3.2.2 shows the PAM-4 driver using N-over-N architecture. This structure

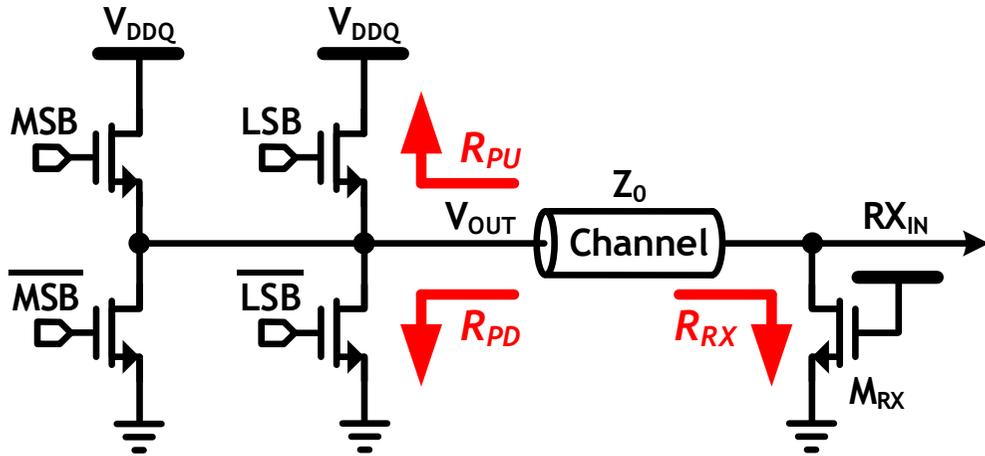


Figure 3.2.2. Voltage-mode PAM-4 driver using N-over-N structure.

consists of two pull-up drivers and two pull-down drivers. In order for the driver to satisfy impedance matching, the impedance of the transmitter seen from the channel must equal the characteristic impedance of the channel, which can be expressed as follows:

$$R_{PU} \parallel R_{PD} = Z_0 \quad (3.2.1)$$

where R_{PU} is the impedance sum of the turned on pull-up drivers, R_{PD} is the impedance sum of the turned on pull-down drivers, and Z_0 is the characteristic impedance of the channel. The level of the transmitter output is determined by the ratio of the impedance of each part of the interface, which can be expressed as follows:

$$\frac{R_{PD} \parallel R_{RX}}{R_{PU} + R_{PD} \parallel R_{RX}} \cdot V_{DDQ} = V_{OUT}, \quad (3.2.2)$$

where R_{RX} is the termination impedance of the receiver. As described in Chapter 2.2.1, the generated four levels must have the same spacing for high linearity. To satisfy impedance matching and high linearity at the same time, the pull-up driver to which the MSB signal is applied as input and the pull-down driver to which the inverted MSB signal enters must have an impedance of $1.5 \cdot Z_0$ when turned on. Pull-up/down drivers to which the LSB and inverted LSB are applied as inputs must have an impedance of $3 \cdot Z_0$, and receiver termination must have an impedance of Z_0 .

Figure 3.2.3 shows the operation of the single-ended voltage-mode PAM-4 driver using N-over-N architecture. To make the highest level, MSB and LSB turn on two pull-up drivers. To generate middle levels, the pull-up driver and pull-down driver are turned on alternately. Finally, when only two pull-down drivers are turned on, the lowest level is generated. By these operation, the driver can generate four levels, $1/2 \cdot V_{DDQ}$, $1/3 \cdot V_{DDQ}$, $1/6 \cdot V_{DDQ}$, and V_{SSQ} . However, these results assume that the MOSFETs are ideal. The pull-up NMOS constituting the N-over-N driver operates in the saturation region, and the impedance of this pull-up driver can be expressed as follows:

$$R_{PU} \approx \frac{V_{DDQ} - V_{OUT}}{1/2 \mu_n C_{ox} \left(\frac{W}{L}\right) (V_{DDQ} - V_{OUT} - V_{TH})^2 (1 + \lambda(V_{DDQ} - V_{OUT}))}. \quad (3.2.3)$$

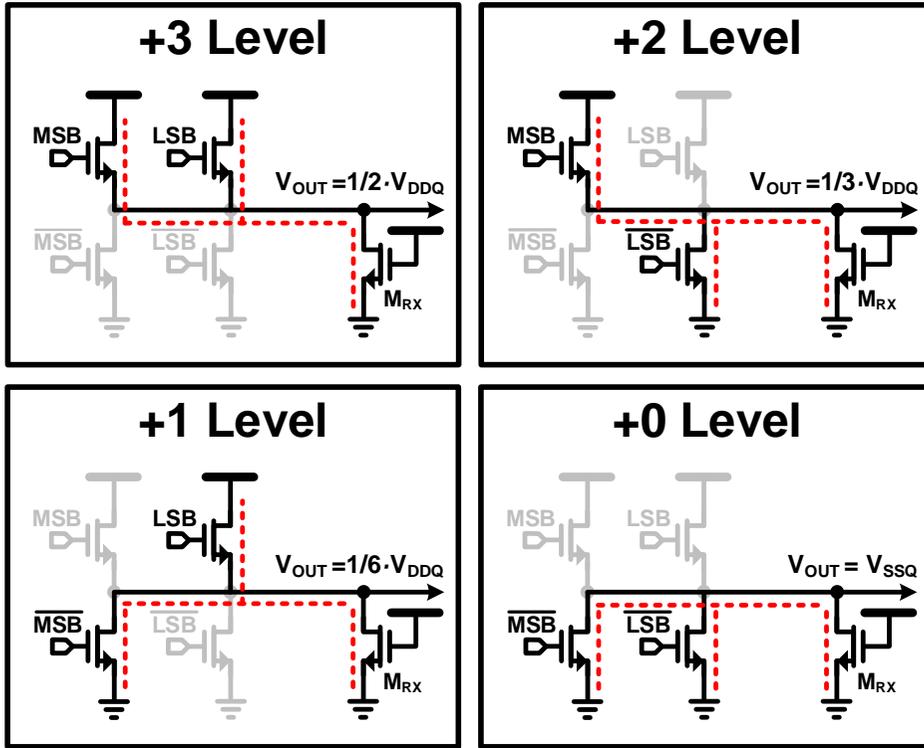


Figure 3.2.3. Operation of generating four levels for PAM-4 driver using N-over-N structure, assuming ideal MOSFETs are used.

On the other hand, the pull-down NMOS of the N-over-N driver operates in a linear region, and the impedance of this pull-down driver can be expressed as follows:

$$R_{PD} \approx \frac{1}{1/2\mu_n C_{ox} \left(\frac{W}{L}\right) (2(V_{DDQ} - V_{TH}) + V_{OUT})}. \quad (3.2.4)$$

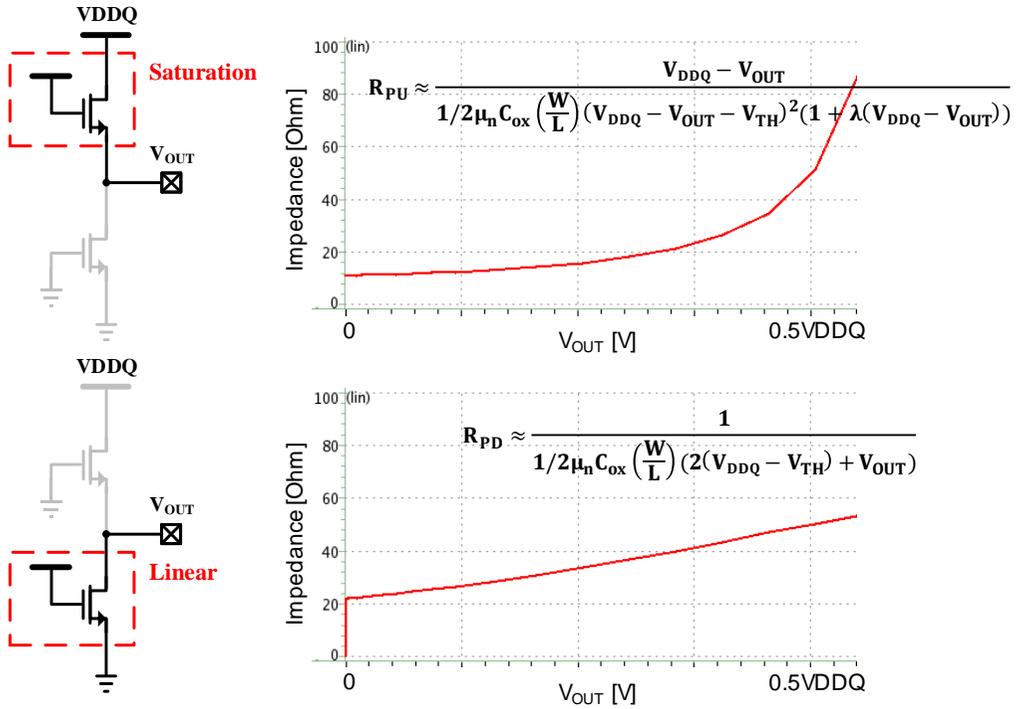
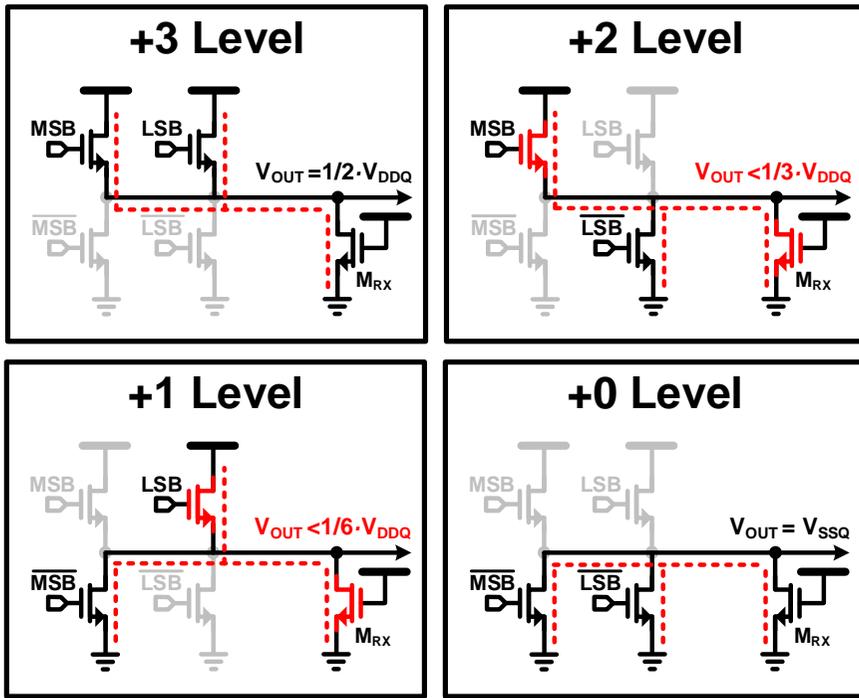


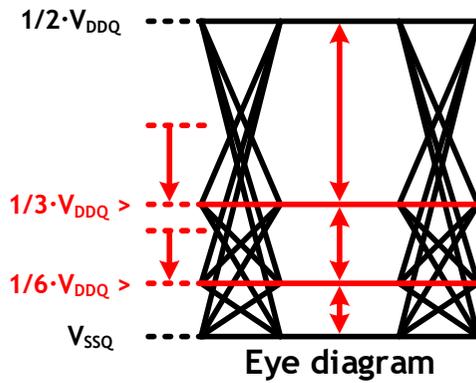
Figure 3.2.4. Impedance of N-over-N driver according to output voltage.

These two equations show that the impedances of the MOSFETs change as V_{OUT} changes. Figure 3.2.4 shows the impedance of the N-over-N driver according to V_{OUT} . This means that the impedances of the pull-up drivers and pull-down drivers change according to the four levels of the PAM-4, which makes the output non-linear. Figure 3.2.5 shows the operation of single-ended voltage-mode PAM-4 driver using N-over-N structure and eye diagram of its output when using real MOSFETs. This assumes that the impedances of all

pull-up / down drivers are set to the desired value at the highest level to maintain the maximum swing level of the output. At middle levels, the impedances of the pull-up / down drivers are changed. Moreover, the memory uses MOSFET for termination of the receiver, and its impedance also changes. Therefore, the level of the output determined by the ratio of impedances deviates from the desired level causing a non-linear output. Impedance variation also causes impedance mismatch with the channel and makes the signal suffer from reflection.



(a)



(b)

Figure 3.2.5. (a) Operation of PAM-4 driver using N-over-N structure, and (b) its output eye diagram with impedance variation.

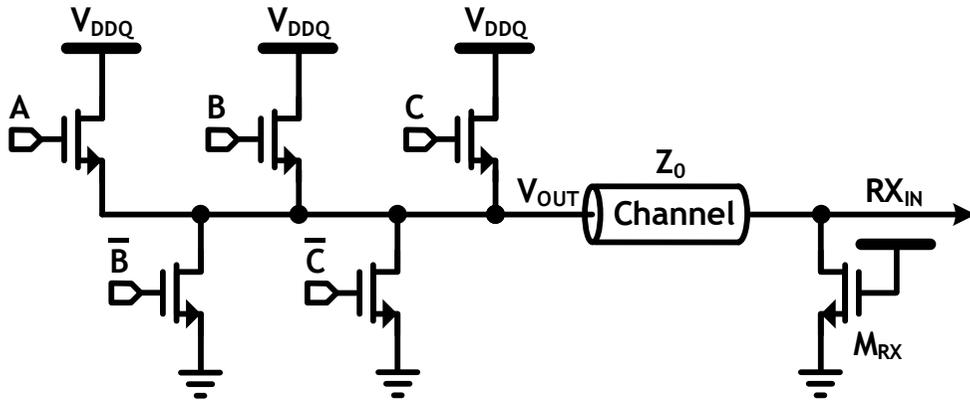


Figure 3.2.6. Proposed impedance-matched PAM-4 driver.

To mitigate the above-mentioned problems, an IM-PAM-4 driver is proposed, as shown in Figure 3.2.6. The driver has three pull-up drivers and two pull-down drivers. Compared to the previous PAM-4 drivers that have two pull-up drivers and two pull-down drivers, the proposed driver has an additional pull-up driver. Previous voltage-mode PAM-4 drivers use series resistors for impedance matching and inductors for high-speed output. However, these make the driver occupy a large area. The proposed driver has a small area with a resistorless and inductorless structure. In addition, the driver adopts single-ended signaling, which uses fewer pins than differential signaling, leading to a small cost. The proposed driver consumes small power by using a voltage-mode topology that consumes less power than the current-mode driver topology. All of these advantages make this driver suitable for memory applications.

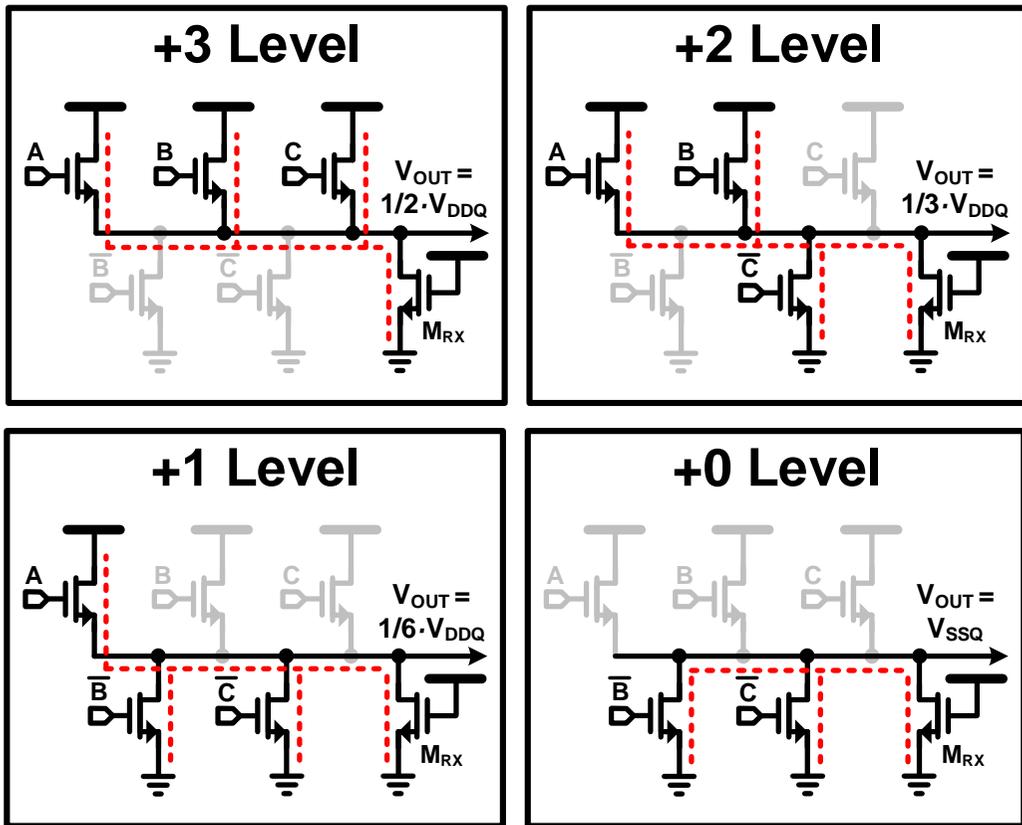


Figure 3.2.7. Operation of proposed IM-PAM-4 driver.

Figure 3.2.7 shows the operation that the proposed driver creates four levels. Three pull-up drivers are turned on to create the highest level, and the number of pull-up drivers turned on is reduced by one to make the lower levels. In order to satisfy impedance matching, Eq. (3.2.1) must be satisfied, and for linearity, the intervals of output levels must be the same. Compared to the previous drivers, the additional pull-up driver increases the

number of output cases of the proposed driver, which makes this driver satisfy impedance matching and high linearity for all levels of the PAM-4. For each level, the impedance of each branch can be obtained through Eq. (3.2.1) and Eq. (3.2.2), and the results can be expressed as follows:

$$R_{MU0} \parallel R_{MU1} \parallel R_{MU2} = Z_0, \quad (3.2.5)$$

$$R_{MU0} \parallel R_{MU1} = \frac{3Z_0 \cdot (Z_0 - \Delta_0)}{2Z_0 - \Delta_0}, \quad (3.2.6)$$

$$R_{MU0} = \frac{6Z_0 \cdot (Z_0 - \Delta_1)}{2Z_0 - \Delta_1}, \quad (3.2.7)$$

$$R_{MD0} \parallel R_{MD1} = \frac{6Z_0 \cdot (Z_0 - \Delta_1)}{4Z_0 - 5\Delta_1}, \quad (3.2.8)$$

$$R_{MD1} = \frac{3Z_0 \cdot (Z_0 - \Delta_0)}{Z_0 - 2\Delta_0}, \quad (3.2.9)$$

where MU0, MU1 and MU2 represent three pull-up drivers, and MD0 and MD1 represent two pull-down drivers. Δ_0 represents the impedance variation of receiver termination at “+2 level”, Δ_1 represents the impedance variation of receiver termination at “+1 level”. As mentioned earlier, the impedance variation of receiver termination must also be considered in order to obtain a high linearity output. The proposed driver satisfies both impedance matching and high linearity by considering both the impedance variation of each branch and the receiver termination. Although the pull-up driver has been increased one more, the total impedance of the pull-up drivers is the same Z_0 as the previous driver, which means that the total size of the pull-up drivers is the same, as shown in Eq. (3.2.5).

Figure 3.2.8 shows the circuit implementation of the proposed IM-PAM-4 driver. Each pull-up/down driver is segmented into a binary 5-bit. This segmentation allows the size of each branch to be adjusted independently.

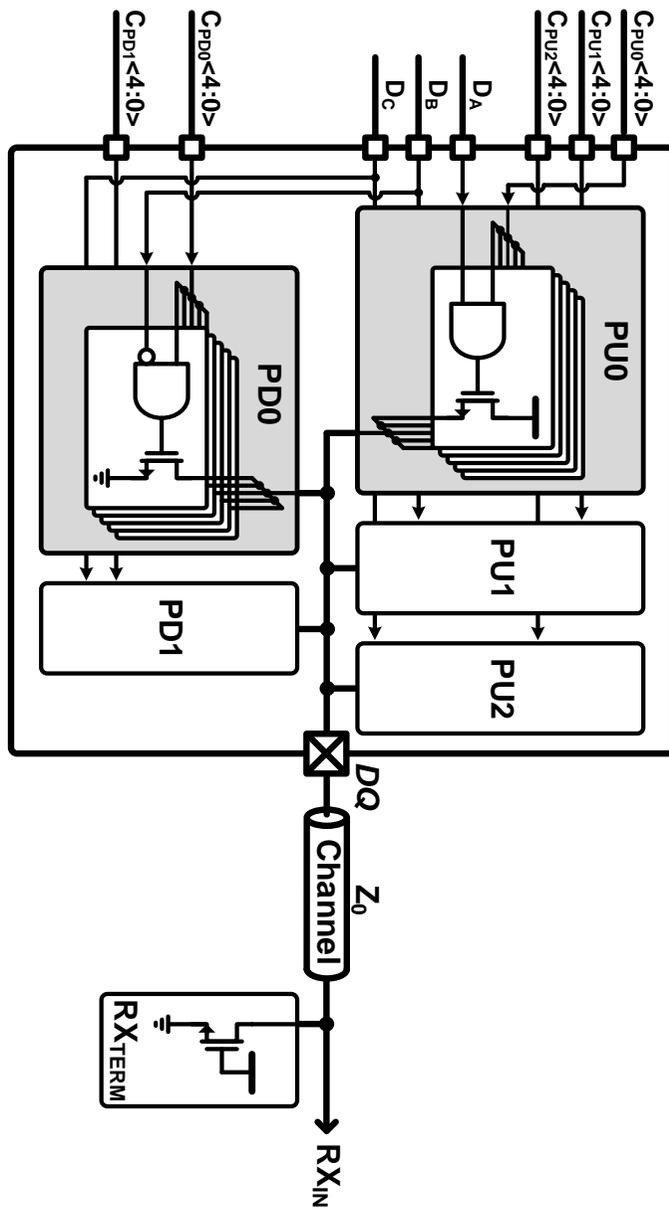
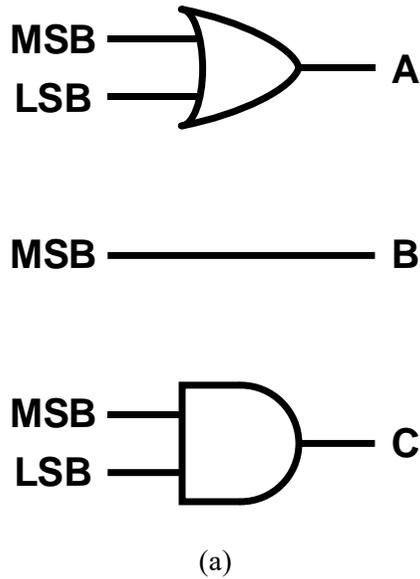


Figure 3.2.8. Circuit implementation of proposed IM-PAM-4 driver.



Encoder Logic
• $A = \text{MSB} + \text{LSB}$
• $B = \text{MSB}$
• $C = \text{MSB} \cdot \text{LSB}$

(b)

Figure 3.2.9. (a) Block diagram and (b) encoding logic of encoder for IM-PAM-4 driver.

The IM-PAM-4 driver does not use the MSB and LSB signals used in previous drivers, which means that a separate encoder is required. As can be seen from the operation of the IM-PAM-4 driver in Figure 3.2.7 that creates four levels, the encoding logic for the IM-PAM-4 driver is simple. Figure 3.2.9 shows the block diagram and encoding logic of the

encoder for the proposed driver. The structure of the encoder for each signal consists of at most one logic gate, which does not burden the entire system. In addition, the encoder is located in front of the 4: 1 serializer, so the encoder operates at low speed, and the jitter generated from the encoder is compensated by the sampling clock in the 4: 1 serializer, as shown in Figure 3.1.1.

3.3 3-POINT ZQ CALIBRATION FOR PAM-4

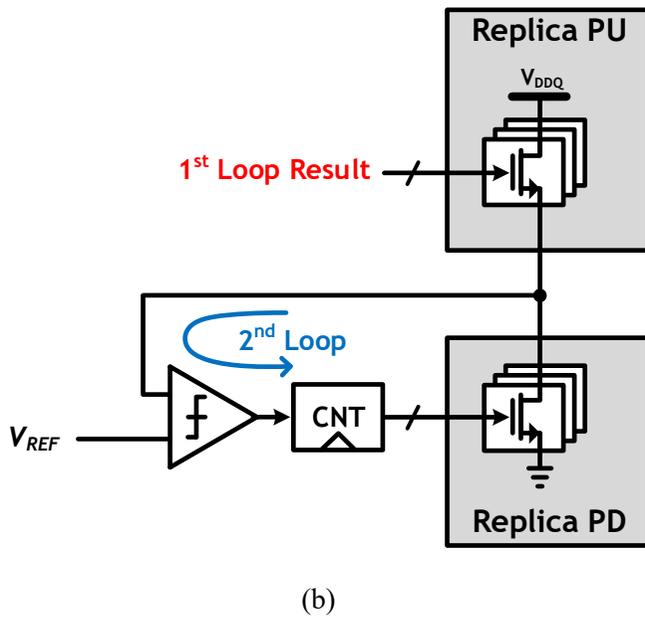
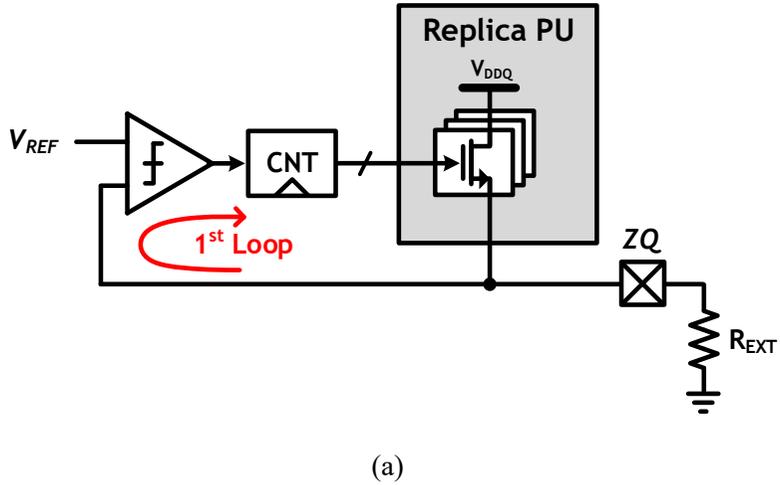


Figure 3.3.1. Conventional ZQ calibrations of (a) pull-up driver and (b) pull-down driver for NRZ signaling.

The output impedance of a voltage-mode driver determines the signal amplitude and prevents reflection through impedance matching with the channel. Therefore, it is important to have the accurate output impedance. ZQ calibration is a circuit used for impedance calibration in memory [1.1.5]. Figure 3.3.1 shows conventional ZQ calibrations for NRZ signaling. This circuit consists of two blocks. The first block is a block that calibrates the impedance of the pull-up driver through an external reference resistor, and the second block is a block that calibrates the impedance of the pull-down driver using the matched pull-up driver as a reference. Since this circuit is for NRZ signaling, only one pull-up driver and one pull-down driver can be calibrated. Therefore, a new ZQ calibration for PAM-4 signaling is required.

Circuits that control the impedance of previous PAM-4 drivers operate only at one point [3.3.1]. these cannot compensate for the variation in the impedance of the driver as a result of changing the output level, mentioned in Chapter 3.2. To ensure that the driver has the correct impedance at all PAM-4 levels, calibration should be performed at all output levels.

Figure 3.3.2 shows the overall architecture of the proposed 3-point ZQ calibration for PAM-4 signaling. The proposed ZQ calibration consists of two blocks. The first block uses an external reference resistor just like the ZQ calibration for conventional NRZ signaling. The second block actually contains the DQ driver used for the interface. As mentioned in Chapter 3.2, the driver needs to know the impedance of the receiver termination in order to satisfy impedance matching and high linearity at the same time. The second block is connected to the receiver termination through the DQ pin, which means that the proposed

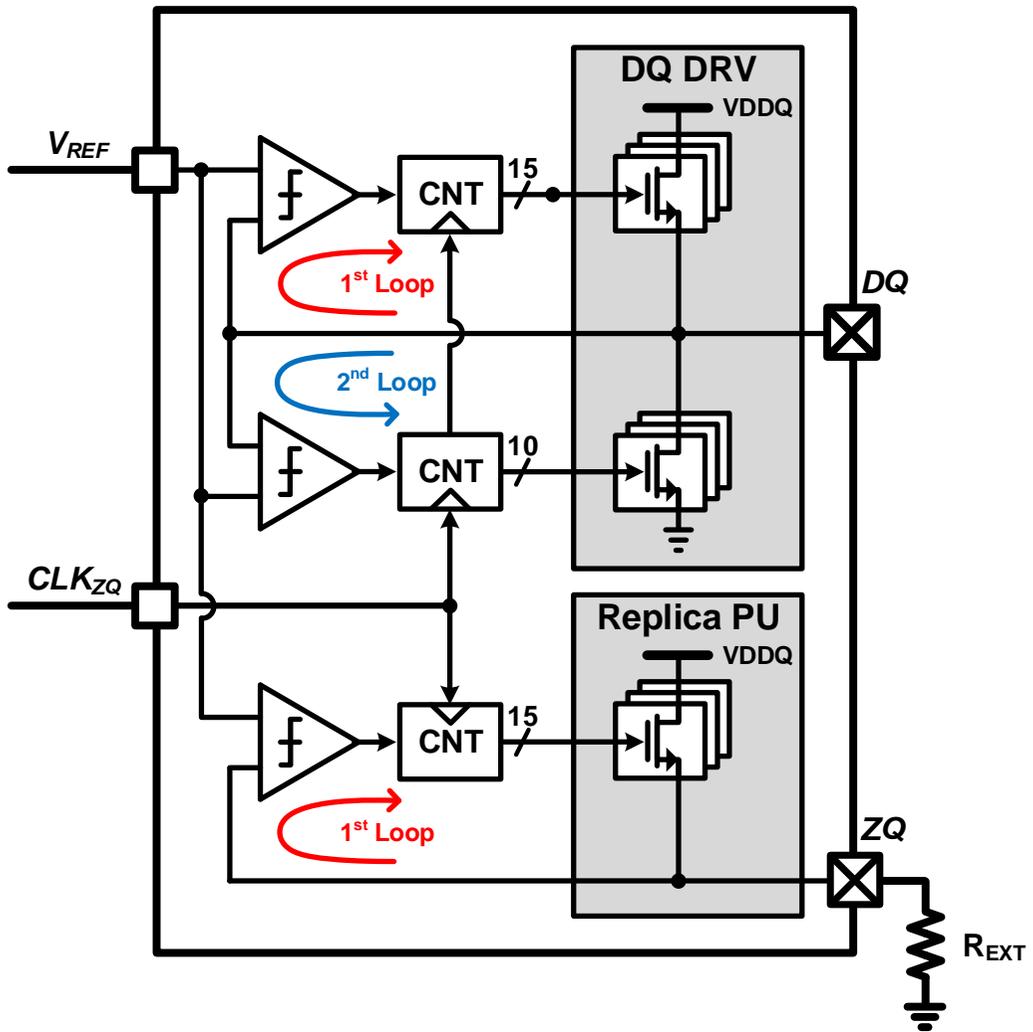
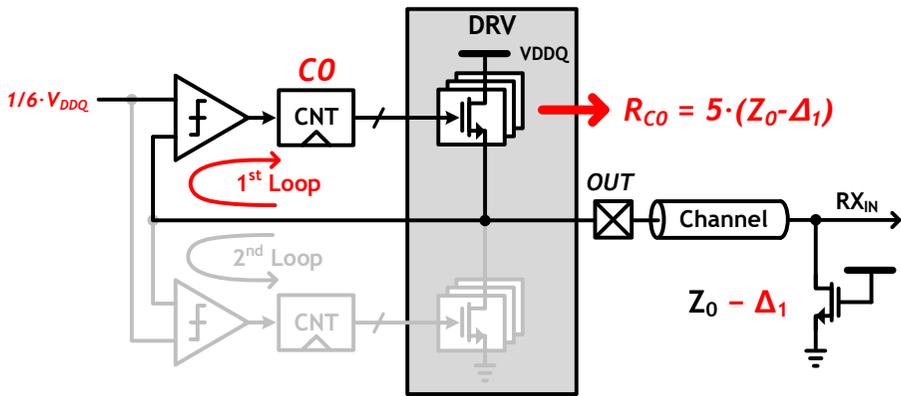
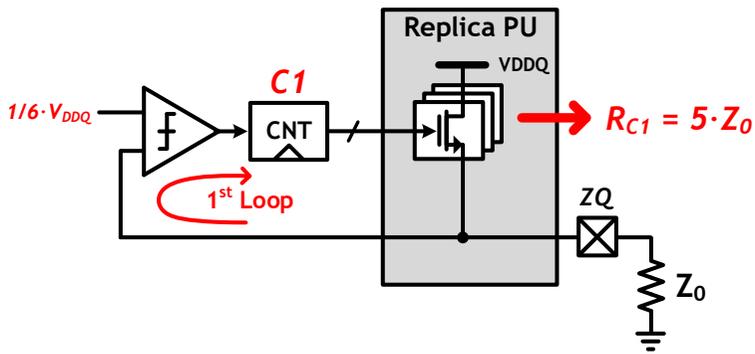


Figure 3.3.2. Overall architecture of proposed 3-point ZQ calibration.



(a)



(b)

Figure 3.3.3. First loop operation of proposed ZQ calibration at $1/6 \cdot V_{DDQ}$.

ZQ calibration takes the impedance information of the receiver termination. Moreover, unlike previous PAM-4 impedance calibration loops, which only calibrate at one point, the proposed calibration has three calibration points. These three calibration points can accurately compensate for the changed impedance at each level of the PAM-4, allowing the driver to accurately output all levels.

Figure 3.3.3 shows the first loop operation at $1/6 \cdot V_{DDQ}$, one of the three calibrating points. The two blocks extract impedance information of the external reference resistor and the receiver termination at the target level. These operations assume that the termination of the receiver is set to impedance matching at the highest level of the PAM-4 to maintain the maximum output level. As shown in Figure 3.3.3 (a), the receiver termination has the impedance variation at the target level, $1/6 \cdot V_{DDQ}$, and the pull-up driver has the impedance based on the receiver termination. On the other hand, as shown in Figure 3.3.3 (b), another block gets accurate impedance information from the external reference resistor. The calibration results obtained by these operations, C0 and C1, can be expressed as follows:

$$R_{C0} = 5 \cdot (Z_0 - \Delta_1), \quad (3.3.1)$$

$$R_{C1} = 5 \cdot Z_0, \quad (3.3.2)$$

Where Δ_1 indicates the impedance variation of receiver termination at “+1 level” as in Chapter 3.2. Derived from these two equations, the ZQ code value corresponding to the pull-up driver’s impedance value of Eq. (3.2.7) can be obtained, which can be expressed as follows:

$$R_{MU0} = \frac{6Z_0 \cdot (Z_0 - \Delta_1)}{2Z_0 - \Delta_1} = 6(Z_0 \parallel (Z_0 - \Delta_1)) = \frac{6}{5} \cdot R_{C0+C1} = R_{5/6 \cdot (C0+C1)}. \quad (3.3.3)$$

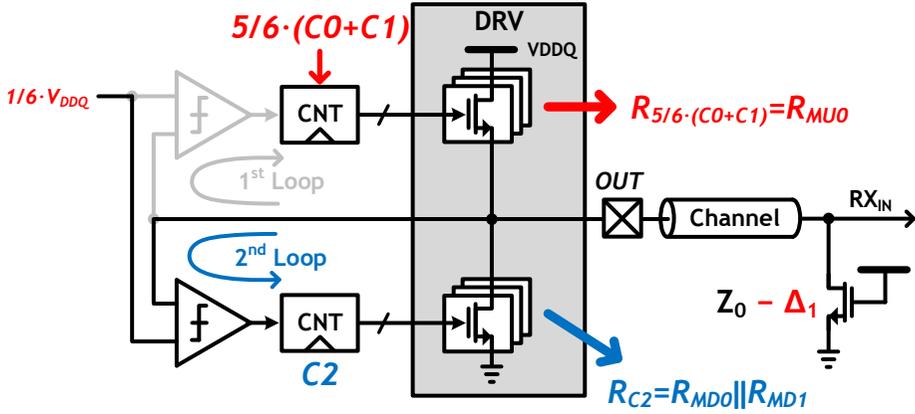
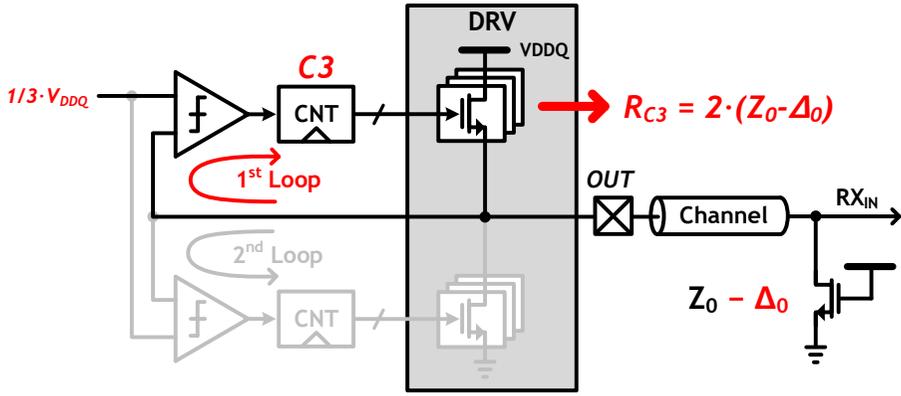


Figure 3.3.4. Second loop operation of proposed ZQ calibration at $1/6 \cdot V_{DDQ}$.

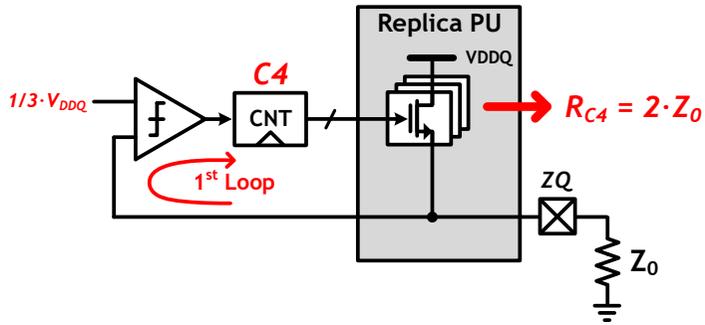
Figure 3.3.4 shows the second loop's operation at $1/6 \cdot V_{DDQ}$. This loop calibrates the pull-down driver while applying the result of the first loop, Eq. (3.3.3), to the pull-up driver. This operation accurately obtains the impedance value of the pull-down driver to have at the target level. Therefore, the result of this operation, $C2$, has the same impedance as Eq. (3.2.8), which can be expressed as follows:

$$R_{MD0} \parallel R_{MD1} = \frac{6Z_0 \cdot (Z_0 - \Delta_1)}{4Z_0 - 5\Delta_1} = R_{C2}. \quad (3.3.4)$$

Figure 3.3.5 and Figure 3.3.6 show operation of the ZQ calibration at $1/3 \cdot V_{DDQ}$, which is the second point of three calibration points. This operation is similar to the operation at the



(a)



(b)

Figure 3.3.5. First loop operation of proposed ZQ calibration at $1/3 \cdot V_{DDQ}$.

first point, $1/6 \cdot V_{DDQ}$. The first loop's results are C3 and C4, which can be expressed as follows:

$$R_{C3} = 2 \cdot (Z_0 - \Delta_0), \quad (3.3.5)$$

$$R_{C4} = 2 \cdot Z_0, \quad (3.3.6)$$

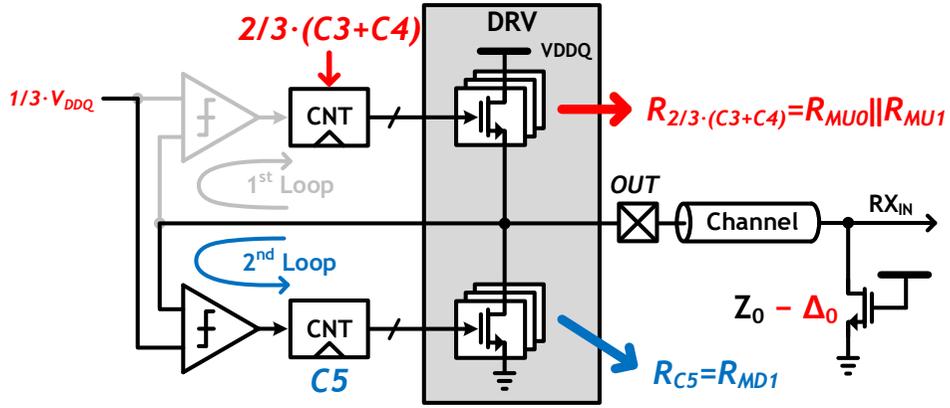


Figure 3.3.6. Second loop operation of proposed ZQ calibration at $1/3 \cdot V_{DDQ}$.

Where Δ_0 indicates the impedance variation of receiver termination at “+2 level” as in Chapter 3.2. These equations show the impedance values of pull-up drivers at target level, $1/3 \cdot V_{DDQ}$. From these two equations, the ZQ code corresponding to the impedance of Eq. (3.2.6) can be obtained, which can be expressed as follows:

$$\begin{aligned}
 R_{MU0} \parallel R_{MU1} &= \frac{3Z_0 \cdot (Z_0 - \Delta_0)}{2Z_0 - \Delta_0} = 3(Z_0 \parallel (Z_0 - \Delta_0)) \\
 &= \frac{3}{2} \cdot R_{C3+C4} = R_{2/3 \cdot (C3+C4)}. \tag{3.3.7}
 \end{aligned}$$

The second loop works with this result, Eq. (3.3.7), and accurately obtains the impedance value of the pull-down driver to have at the target level. Therefore, the result of this

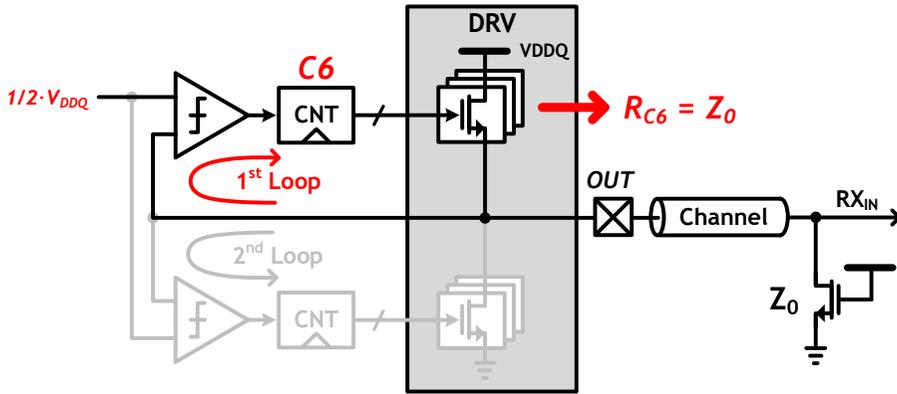


Figure 3.3.7. First loop operation of proposed ZQ calibration at $1/2 \cdot V_{DDQ}$.

operation, C5, has the same impedance as Eq. (3.2.9), which can be expressed as follows:

$$R_{MD1} = \frac{3Z_0 \cdot (Z_0 - \Delta_0)}{Z_0 - 2\Delta_0} = R_{C5}. \quad (3.3.8)$$

Figure 3.3.7 shows operation of the ZQ calibration at $1/2 \cdot V_{DDQ}$, which is the third point of three calibration points. This operation, unlike the operation at other points, operates only the first loop in one block, because the calibration previously assumed that the receiver termination has the impedance of Z_0 at the highest level, $1/2 \cdot V_{DDQ}$. Therefore, the result of this operation, C6, can be expressed as follows:

$$R_{C6} = Z_0. \quad (3.3.9)$$

From this result, the ZQ code corresponding to the impedance of Eq. (3.2.5) can be expressed as follows:

$$R_{MU0} \parallel R_{MU1} \parallel R_{MU2} = Z_0 = R_{C6}. \quad (3.3.10)$$

The equations obtained from the three calibration points indicate the size of each of the proposed three pull-up drivers and two pull-down drivers of the proposed IM-PAM-4 driver, which can be expressed as follows:

$$C_{MU0} = \frac{5}{6}(C0 + C1), \quad (3.3.11)$$

$$C_{MU1} = \frac{2}{3}(C3 + C4) - \frac{5}{6}(C0 + C1), \quad (3.3.12)$$

$$C_{MU2} = C6 - \frac{2}{3}(C3 + C4), \quad (3.3.13)$$

$$C_{MD0} = C2 - C5, \quad (3.3.14)$$

$$C_{MD1} = C5. \quad (3.3.15)$$

These results allow the IM-PAM-4 driver to have impedance matching and high linearity at all output levels, taking into account the impedance variation of both the driver and the receiver.

CHAPTER 4

REFLECTION-BASED DUOBINARY TRANSMITTER

To increase memory density, the stacked die packaging with multiple DRAM die stacked vertically in one package is widely used. However, combined with the center-pad structure, the structure creates stubs that cause short reflections. We have proposed the reflection-based duobinary transmitter to mitigate this problem. The proposed transmitter uses reflection for duobinary signaling. The 2-tap opposite FFE and the slew-rate control are used to increase signal integrity.

4.1 BIDIRECTIONAL DUAL-RANK MEMORY SYSTEM

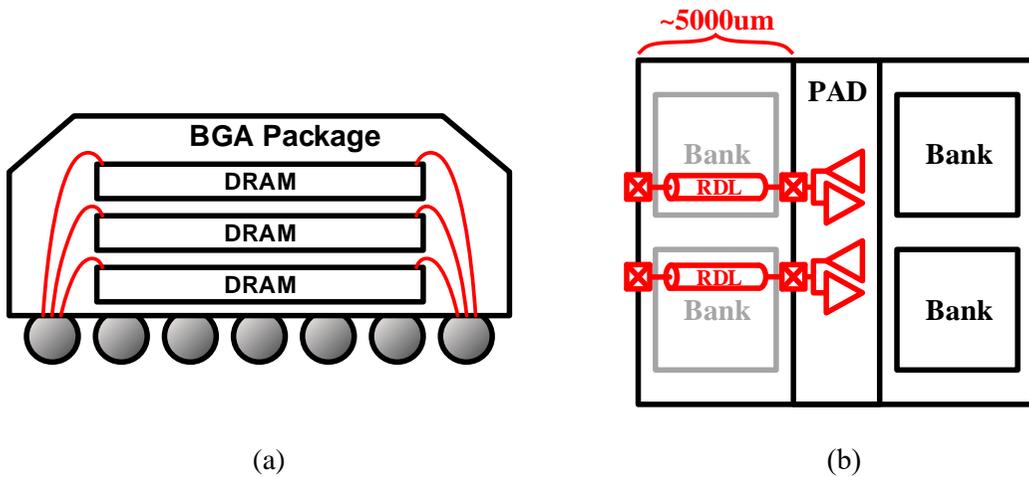


Figure 4.1.1. (a) Stacked-die packaging and (b) center-pad structure for memory.

As mentioned in Chapter 1, one of the parallelization methods to increase density is 3DS memory. Figure 4.1.1 (a) shows the stacked-die packaging of stacking multiple DRAM dies vertically in one package. This technique reduces costs by using multiple dies in one package, and enables high density in one footprint [1.1.3]. In addition, this parallelization method has a faster access timing because the path is shorter than that of other parallelization methods [4.1.1].

As shown in Figure 4.1.1 (b), memory uses a center-pad structure that places I/O PADs at the center of the chip for the efficient placement and routing of banks that occupy most of the area [4.1.2]. In this structure, a redistribution layer (RDL) is used to connect the

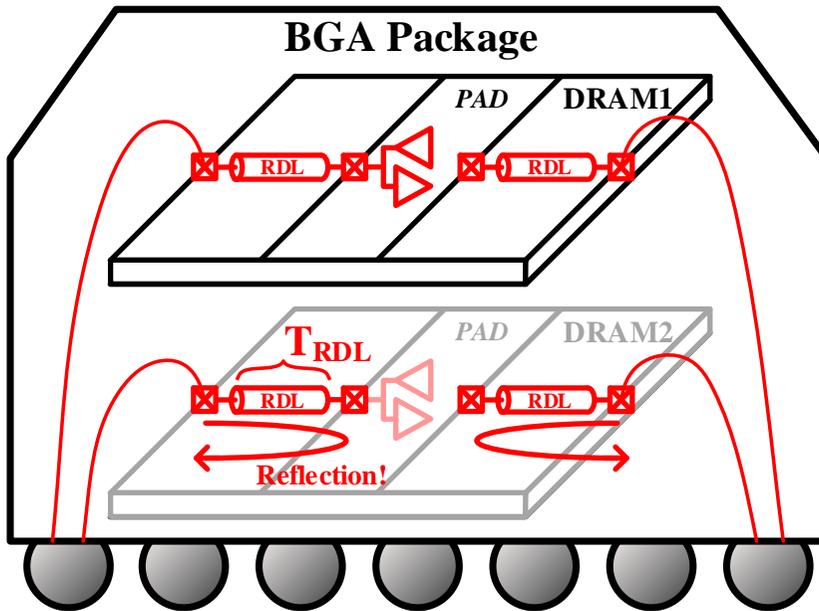


Figure 4.1.2. Overall architecture of stacked die packaging with center-pad structure.

PADs located at the center to the bonding metal at the edge. Figure 4.1.2 shows the overall structure of a memory with a center-pad structure when using stacked die packaging technology. At low speeds, RDL is treated as a lumped model without causing problems. However, as memory bandwidth increases, RDL is interpreted as a transmission line that causes reflection. Therefore, 3DS memory using stacked-die packaging technology is recognized as a multi-drop topology at high speed.

The generated reflection arrives at the receiver after twice T_{RDL} , flight time of RDL, compared to the main signal. Since RDL is a trace inside the chip, the timing at which

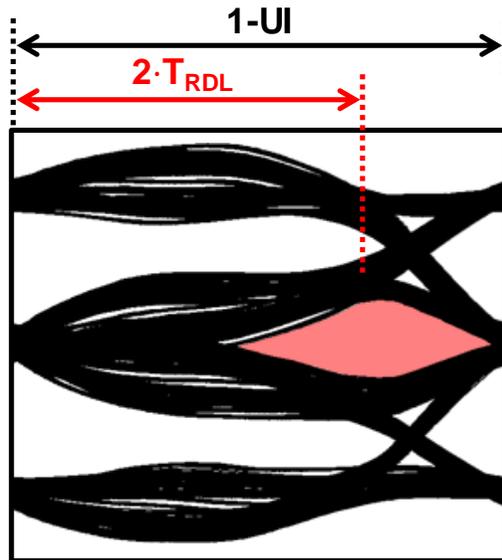
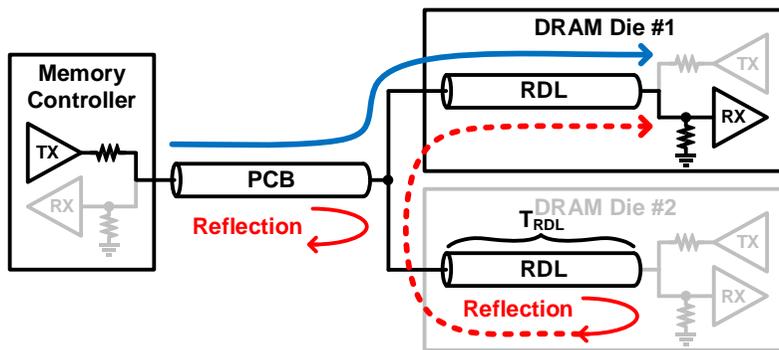
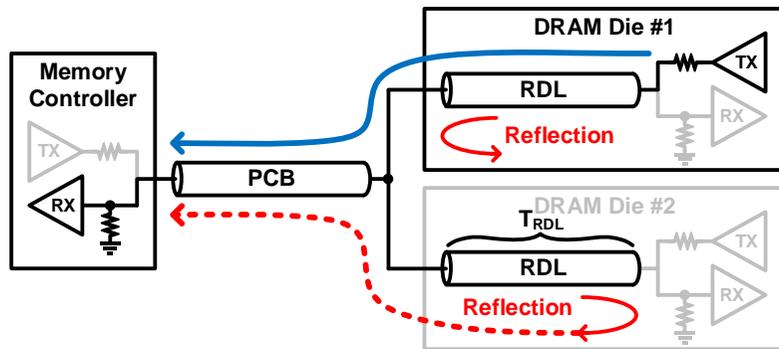


Figure 4.1.3. Simulated eye diagram with 2-rank memory system.

reflection arrives is shorter than the previous multi-drop structures. The length of RDL used for the current memory connecting the center pad and the bonding metal at the edge is about 5000um, and the flight time is about 30ps. This means that the reflection arrives after about 1-UI compared to the main signal, considering the current DRAM data-rate. The short reflection resulting from the length of this short stub is the main difference from the previous multi-drop structures or stub problems. Figure 4.1.3 shows a simulated eye diagram in a 2-rank structure with two die stacked. The eye diagram is distorted due to reflection, and the eye height and eye width are reduced.



(a)



(b)

Figure 4.1.4. (a) READ and (b) WRITE operation of 2-rank memory system.

Memory uses bidirectional signaling to reduce the number of pins. Figure 4.1.4 shows WRITE and READ operation of 2-rank memory. In both operations, a reflection problem due to the unused memory occurs and the reflected signal arrives after $2 \cdot T_{RD_L}$ compared to the main signal.

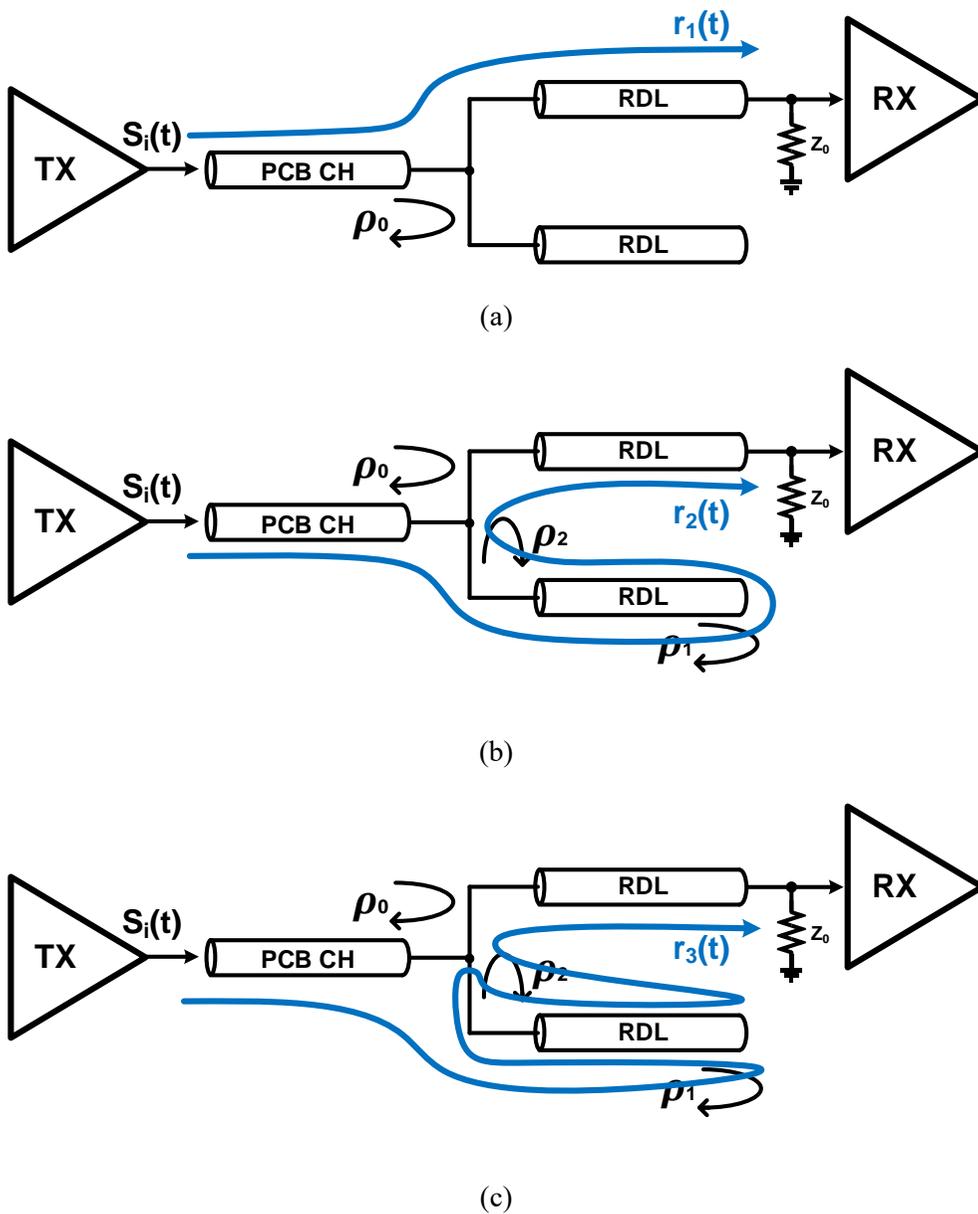


Figure 4.1.5. (a) First, (b) second and (c) third arrived signal.

Figure 4.1.5 shows the signals arriving at the receiver in a 2-rank memory system in order. The first arriving signal passes through the target rank, not reflected at the impedance discontinuity point dividing two ranks. The second arriving signal is reflected from the off DRAM and enters the target DRAM. The third is the signal reflected twice in the RDL stub. Figure 4.1.6 shows the lattice diagram for the signal arriving at the receiver in a 2-rank memory system. The final signal at the receiver is the sum of the infinitely reflected signals, which can be expressed as follows:

$$r(t) = \sum_{n=1}^{\infty} r_n(t). \quad (4.1.1)$$

However, the signal that arrives late has a small amplitude because it passes through reflection and transmission several times. This means that signals arriving late over a certain time have negligibly small amplitude. Therefore, the total received signal can be expressed as follows:

$$r(t) \approx r_1(t) + \rho_0 \cdot \Gamma_2 \cdot r_1(t - 2 \cdot T_{RDL}). \quad (4.1.2)$$

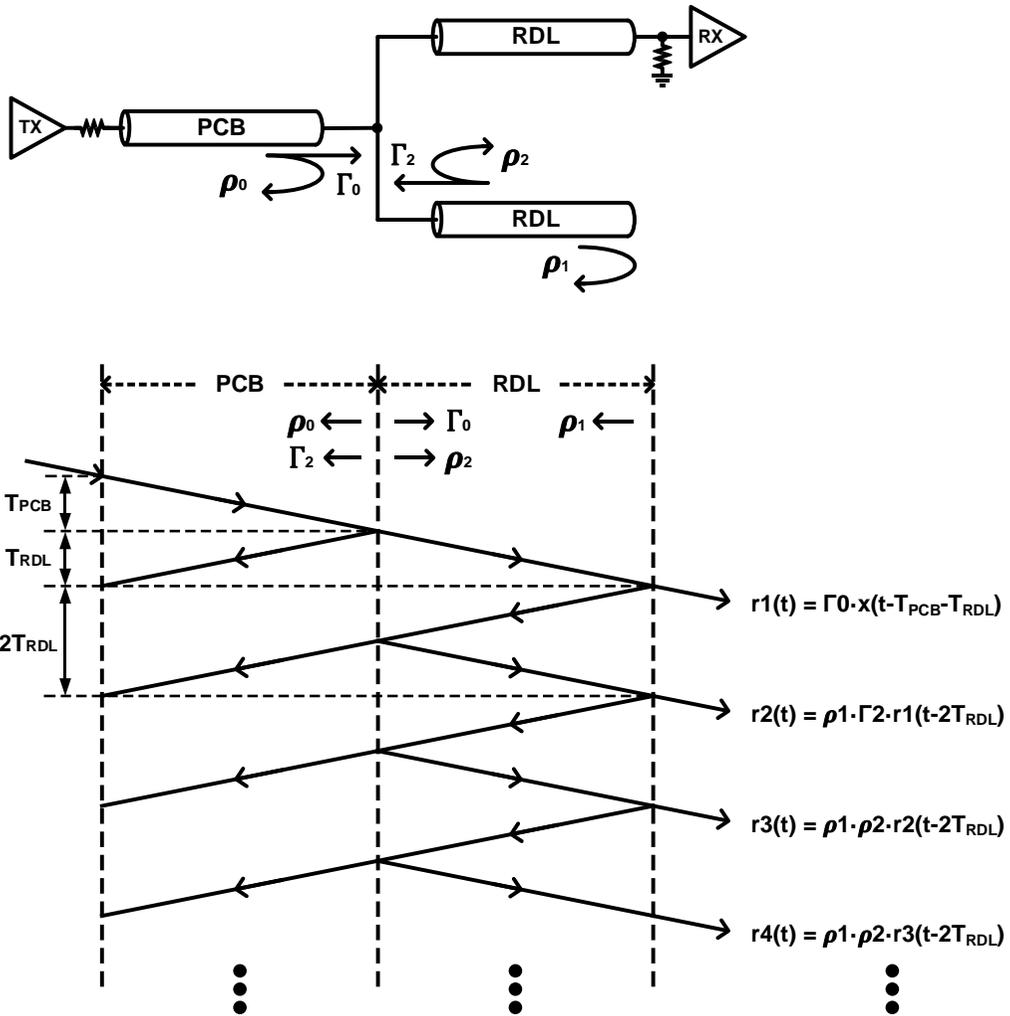


Figure 4.1.6. Lattice diagram for 2-rank memory system.

The received signal consists of two signals, the main signal and the reflected signal. These two signals arrive at the receiver with a timing difference of $2 \cdot T_{\text{RDL}}$. If $2 \cdot T_{\text{RDL}}$ is much longer than $1 \cdot \text{UI}$, the reflected signal can damage the subsequent data and this can be eliminated through equalization, which removes post-cursor ISI like DFE [4.1.3]. If $2 \cdot T_{\text{RDL}}$ is much shorter than $1 \cdot \text{UI}$, it only affects the beginning of the data and is sufficient to ignore. On the other hand, if $2 \cdot T_{\text{RDL}}$ has a length similar to $1 \cdot \text{UI}$, it can no longer be ignored and data is distorted. In the recent multi-rank memory system, $2 \cdot T_{\text{RDL}}$ has a length similar to that of $1 \cdot \text{UI}$ and a solution is needed.

4.2 CONCEPT OF REFLECTION-BASED DUOBINARY SIGNALING

In the multi-drop topology, efforts have been made to remove reflection by structurally satisfying impedance matching. Non-target on-die termination (NT-ODT) prevents reflection from each stub by always turning on on-die termination (ODT) of non-target DRAM in a 2-rank structure [4.2.1]-[4.2.3]. However, current flows through NT-ODT during signaling, which consumes power. This reduces energy efficiency compared to the signal amplitude in the target DRAM. In particular, in a multi-rank structure in which more DRAMs are parallelized, it has lower energy efficiency. Efforts have been made to eliminate impedance discontinuities by using passive resistors on branching nodes in multi-drop buses [4.2.4],[4.2.5]. However, only unidirectional impedance matching is satisfied, and the amplitude of the signal decreases due to the passive resistor. Moreover, like NT-ODT, they have low energy efficiency because they turn on the ODT of unused DRAM.

There have been efforts to eliminate reflections due to stubs with circuits. Decision feedback equalizer (DFE) is applicable when the reflection is arrived after 1-UI due to long stub [4.1.4], but it cannot remove reflection when it has a short stub where reflection arrives near 1-UI. Multi-tone transceiver can transmit data while avoiding the notch frequency generated by reflection [4.2.6]. However, the transceiver requires high speed clock, which is higher than its data-rate.

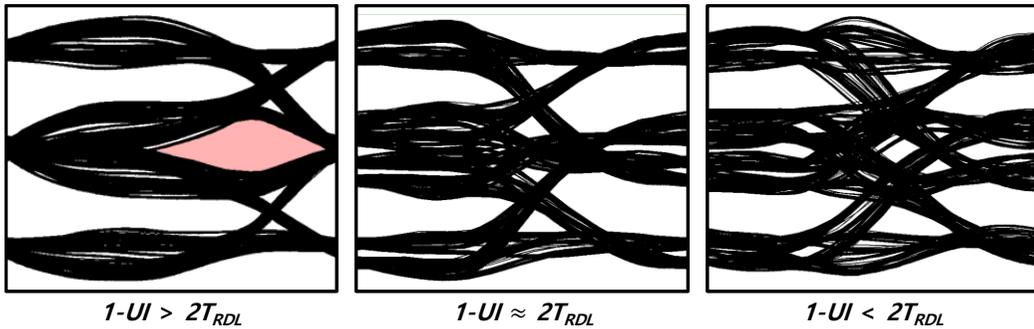


Figure 4.2.1. Simulated eye diagrams at data-rates with 1-UI near $2 \cdot T_{RDL}$ in 2-rank memory system.

As mentioned in Chapter 4.1, the relationship between data-rate and $2 \cdot T_{RDL}$, the time that reflection arrives after the main signal, has a great influence on signal integrity. When the 1-UI is much longer than $2 \cdot T_{RDL}$, reflected signals affect only the front part of the current data, and after that, the amplitudes of reflected signals are sufficiently reduced to not cause a problem. When 1-UI is much shorter than $2 \cdot T_{RDL}$, reflected signals arrive much later than the main signal, so they can be removed by an equalizer such as DFE, which removes the post cursor. On the other hand, when the 1-UI is near $2 \cdot T_{RDL}$, reflection cannot be removed by the existing equalization methods, and reflections reduce signal integrity. Figure 4.2.1 shows eye diagrams obtained from various data-rates where 1-UI is around $2 \cdot T_{RDL}$. When the 1-UI is slightly larger than $2 \cdot T_{RDL}$, the NRZ eye is open but has a small

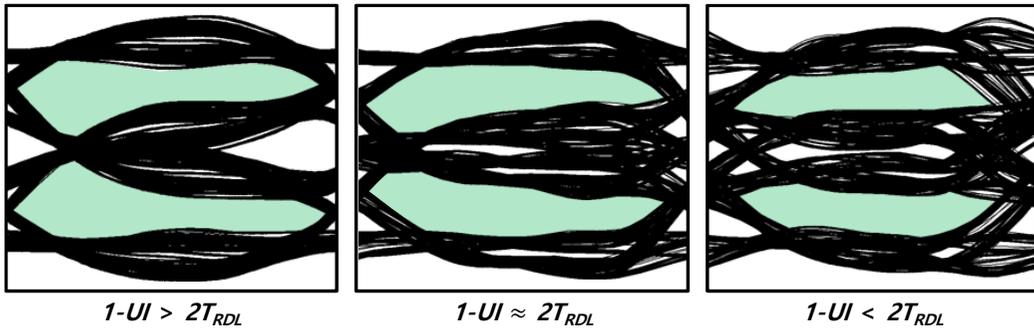


Figure 4.2.2. Transition areas of eye diagrams at data-rates with $1-UI$ near $2 \cdot T_{RDL}$ in 2-rank memory system.

height and width. When $1-UI$ is almost the same as $2 \cdot T_{RDL}$ and $1-UI$ is slightly smaller than $2 \cdot T_{RDL}$, the NRZ eyes are completely closed.

Figure 4.2.2 shows the different areas of the eye diagrams in Figure 4.2.1, which have a wider width and a larger height than the previous eyes. The proposed transmitter considers the areas as eyes of a 3-level signal. The highest level of this 3-level signal means that the current data is "high" and the previous data is also "high" without transition. The lowest level also means that the current data is "low" and the previous is "low" without transition. The middle level means that data is in transition. If the previous data is "low", the current data is "high", and if the previous data is "high", the current data is "low". Therefore, decoding this 3-level signal to an NRZ signal can be expressed as follows:

$$x[n] = r[n] - x[n - 1]. \quad (4.2.1)$$

This signaling method uses reflection unlike previous works to remove reflection. It is energy efficient because it does not require additional power consumption, and it is easy to expand into multi-drop topologies.

The proposed transmitter uses duobinary signaling, which is proved by Eq. (4.2.1) and Eq. (2.3.1) being the same. Previous duobinary transmitters use frequency-dependent channel loss. On the other hand, the proposed transmitter operates using the reflection of a multi-drop system.

As shown in Figure 4.2.2, the duobinary eyes have higher height and wider width than the NRZ eyes, but have slightly distorted shapes. The ideal duobinary eyes have almost half the height of the total signal amplitude and the timing with the maximum height is located near the center of the eye. The length of the RDL that determines T_{RDL} is a given condition and cannot be controlled. Therefore, in order to improve signal integrity, equalizations that control the reflected signals are needed.

4.3 REFLECTION-BASED DUOBINARY TRANSMITTER

4.3.1 OVERALL ARCHITECTURE

Figure 4.3.1.1 shows the overall block diagram of the proposed reflection-based duobinary transmitter. The transmitter has a quarter-rate structure using quadrature clocks, which allows the transmitter to have a low clock frequency. A 32-bit PRBS generator was used for the test. The generated 32-bit PRBS pattern is serialized from low-speed data to high-speed data through a 32:4 serializer. And it is delivered to the driver through a precoder and a 4:1 serializer. 2-tap opposite FFE and slew-rate control are used to achieve high signal integrity.

The signals of the 2-rank memory system are similar to duobinary, but there are two differences, as shown in Eq. (4.1.2) and Eq. (2.3.1). First, the magnitudes of the two signals being added are different. Second, $2 \cdot T_{RDL}$, which is a timing difference between two signals to be added, may not be $1-U$. These two differences make the eye diagram in the 2-rank memory system different from duobinary's ideal eye diagram. Two types of equalizers are used to compensate for these differences, which are 2-tap opposite FFE to compensate for an amplitude difference and slew-rate control to compensate for a timing difference. Since duobinary modulation is basically performed through reflection, the amount compensated by the two equalizers is smaller than the previous equalizers to overcome channel loss.

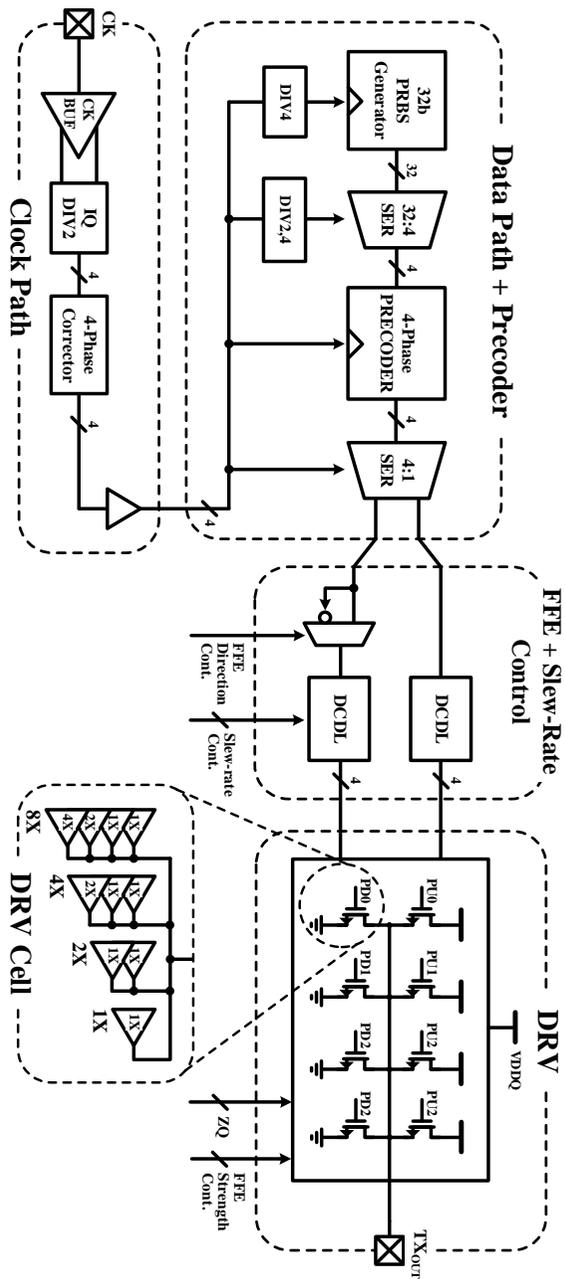


Figure 4.3.1.1. Overall block diagram of proposed reflection-based duobinary transmitter.

4.3.2 EQUALIZATION FOR REFLECTION-BASED DUOBINARY SIGNALING

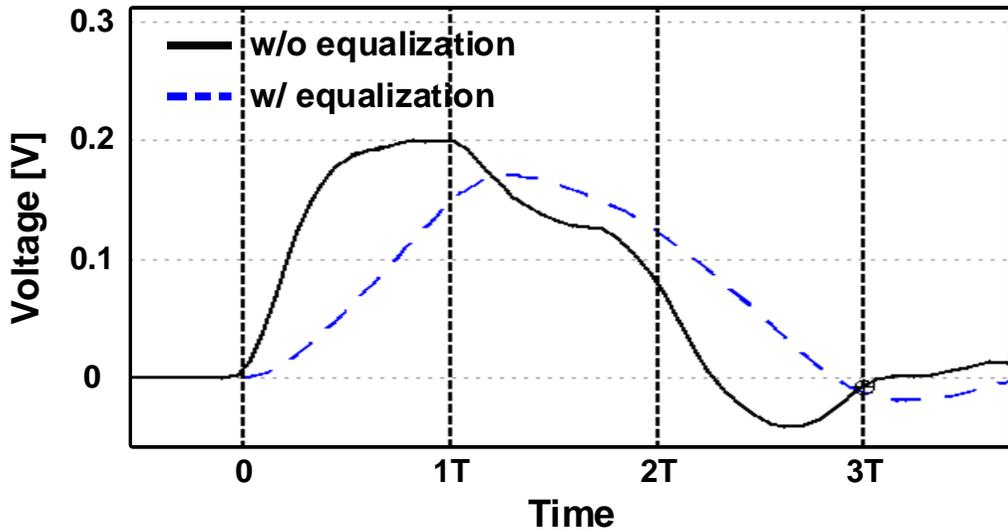
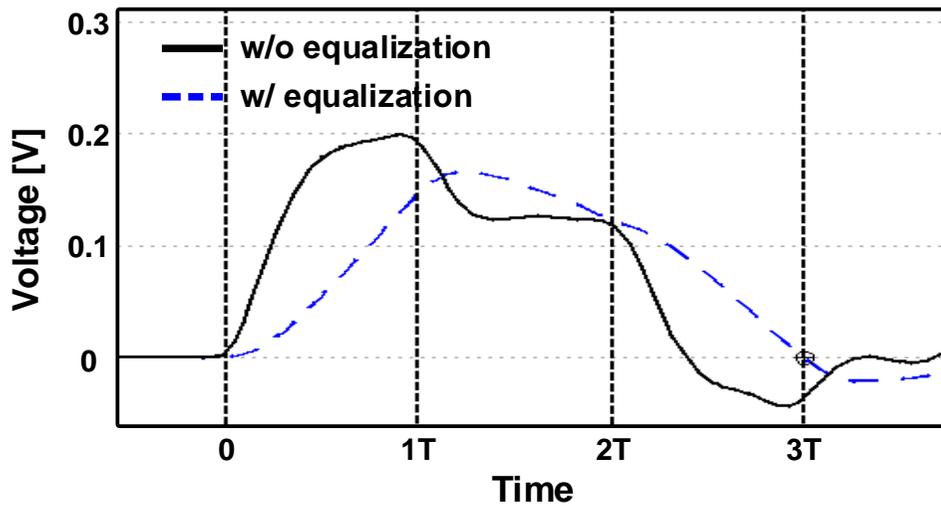


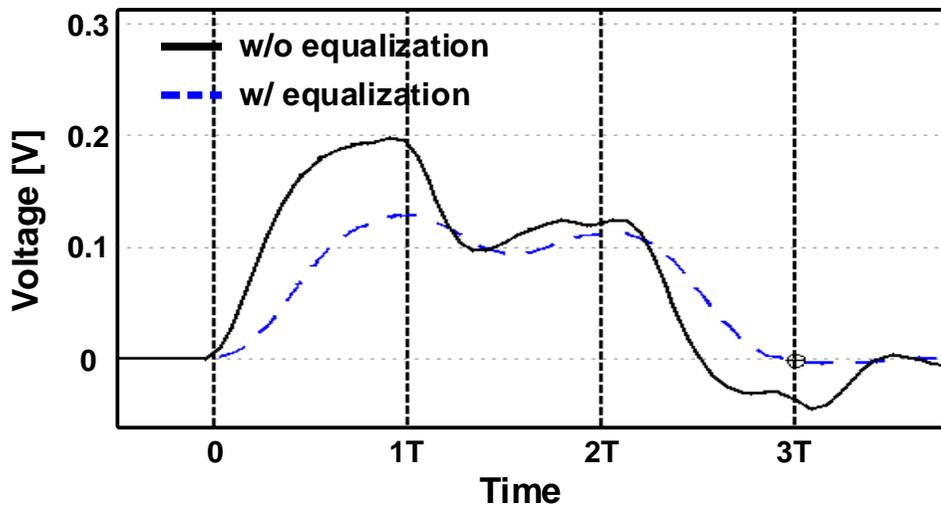
Figure 4.3.1.1. Simulated single-bit response with and without equalization at $1\text{-UI} > 2 \cdot T_{\text{RDL}}$.

From Eq. (4.1.2), the received signal in the 2-rank memory system consists of the main signal and the reflected signal. The reflected signal is delayed by the short stub and added to the main signal. The amplitude of the reflected signal is smaller than the main signal. Figure 4.3.2.1 shows a simulated single-bit response (SBR) at a data rate where 1-UI is longer than $2 \cdot T_{\text{RDL}}$. SBR has a long waveform over about 2-UI due to short reflection. The relatively small amplitude of the reflected signal reaching $2 \cdot T_{\text{RDL}}$ after the main signal makes the SBR have the slow falling edge. The falling edge, which is slower than the rising

edge, causes the phase with the maximum height of the duobinary eye to be located at the front rather than the center, as shown in the left figure of Figure 4.2.2. The slew-rate control is used to make the rising and falling edges have the same slew-rate. The ideal duobinary SBR has the same value across 2-UI , which is half of the signal's full swing range. In the 2-rank memory system, the SBR has the smaller value at $2T$ than that at $1T$ due to the small reflection. A 2-tap opposite FFE is used to compensate for this amplitude difference. The opposite FFE inserts the loss as opposed to previous FFE to compensate for high-frequency loss. Figure 4.3.2.1 also shows the equalized SBR. These two equalization techniques allow the output to have a more ideal duobinary SBR than when there is no equalization. In addition, the amount of equalization is small because a part of the signal is modulated as a duobinary due to reflection, thus they do not place a great burden on the transmitter. Figure 4.3.2.2 shows SBRs at the data rates where 1-UI is similar to $2 \cdot T_{\text{RDL}}$ and 1-UI is smaller than $2 \cdot T_{\text{RDL}}$. SBRs are distorted without equalization, but SBRs are improved through the equalization.



(a)



(b)

Figure 4.3.2.2. Simulated single-bit response with and without equalization at (a) $1\text{-UI} \approx 2 \cdot T_{\text{RDL}}$ and (b) $1\text{-UI} < 2 \cdot T_{\text{RDL}}$.

4.3.3 2D BINARY-SEGMENTED DRIVER

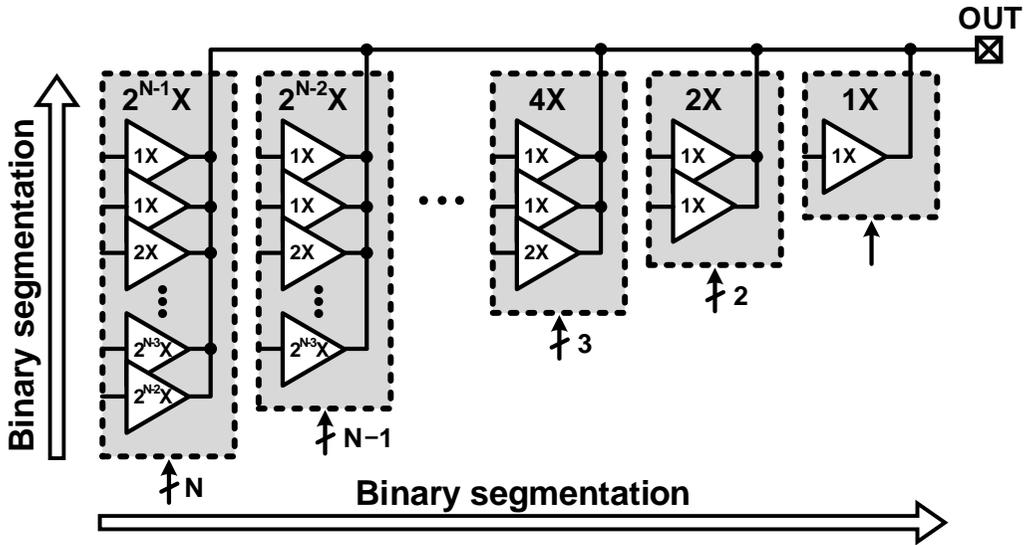


Figure 4.3.3.1. Architecture of 2D binary-segmented driver.

For the 2-tap opposite FFE, the slew-rate control and impedance matching the large number of segmentations is required [4.3.3.1], [4.3.3.2]. An additional driver can be used to reduce the number of segmentations. However, the size of the additional driver must be same with the main driver increasing the C_{I0} . Figure 4.3.3.1 shows the proposed two-dimensional (2D) binary-segmented driver, which reduces the number of segmentations without increasing the overall driver size. The driver is segmented into two-dimensionally. This structure allows the driver to control two characteristic independently.

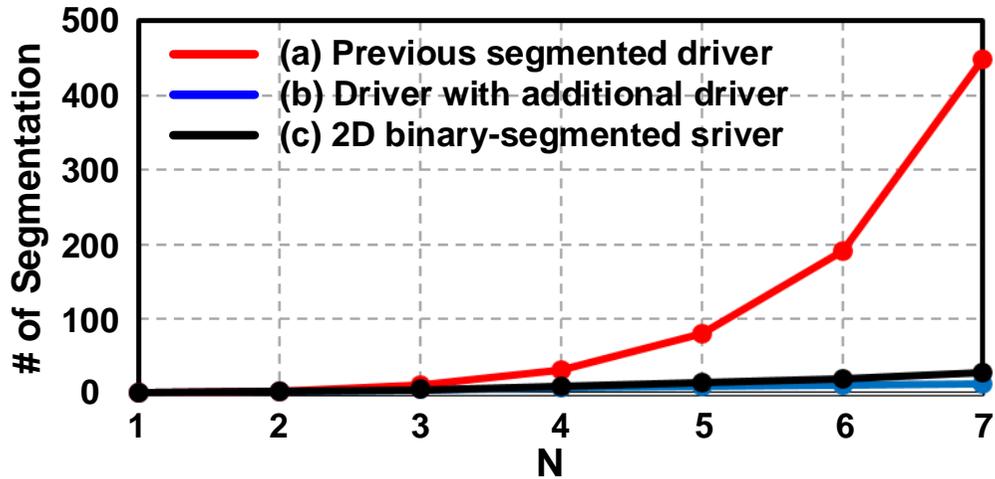


Figure 4.3.3.2. The required number of segmentations to independently control two characteristics.

Figure 4.3.3.2 shows the required number of segmentation for independently control two characteristic of a driver. To control each characteristic with N-bit, the previous segmented driver needs $N \cdot 2^{N-1}$ segments. The driver with additional driver needs only $2 \cdot N$ segments, but occupies large area increasing C_{IO} . On the other hand, the 2D binary-segmented driver needs $N \cdot (N+1)/2$ segments without increasing C_{IO} . Therefore, the proposed driver allows the transmitter to control several output characteristics with the small number of segmentations while maintaining overall driver size.

CHAPTER 5

EXPERIMENTAL RESULTS

5.1 HIGH-LINEARITY AND IMPEDANCE-MATCHED PAM-4 TRANSMITTER

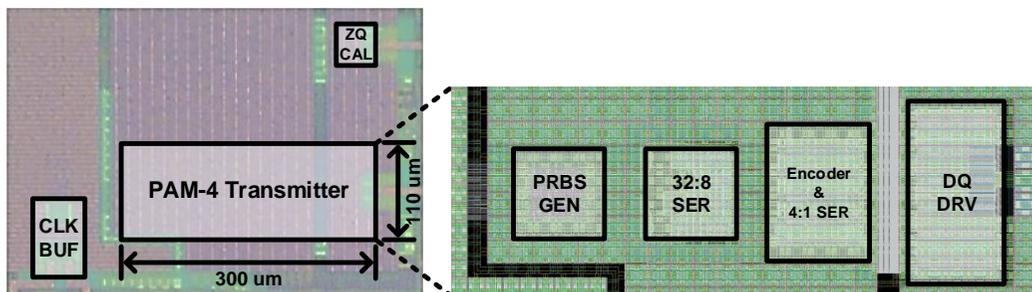


Figure 5.1.1. Chip micrograph of the proposed PAM-4 transmitter.

The prototype chip of proposed PAM-4 transmitter is fabricated in 65nm CMOS process. Figure 5.1.1 shows the micrograph and the total area of the proposed transmitter is 0.0333mm^2 .

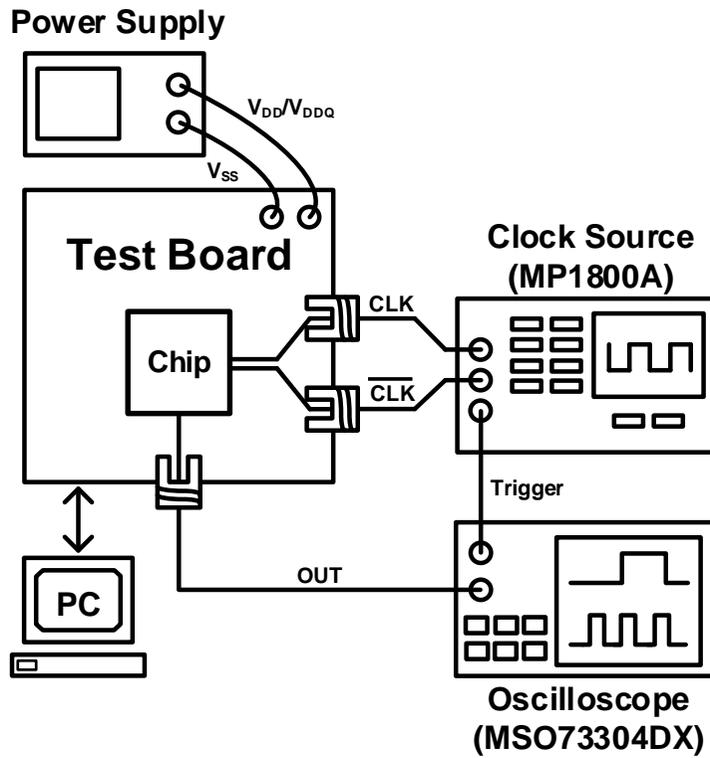
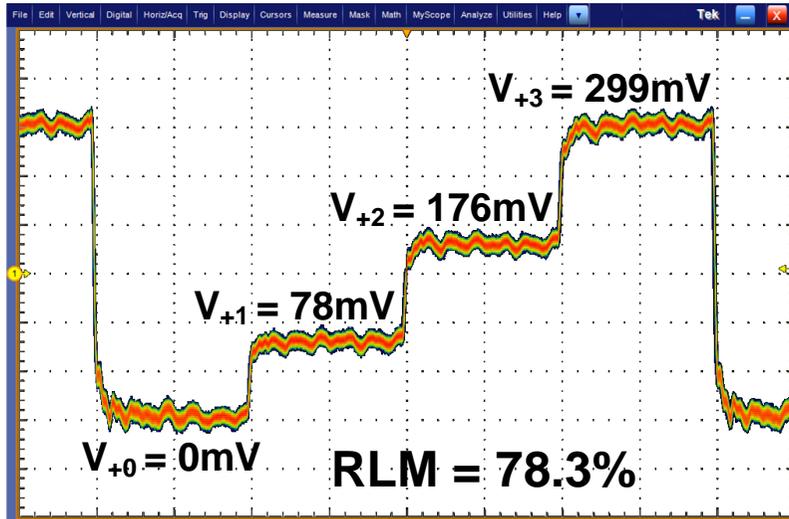
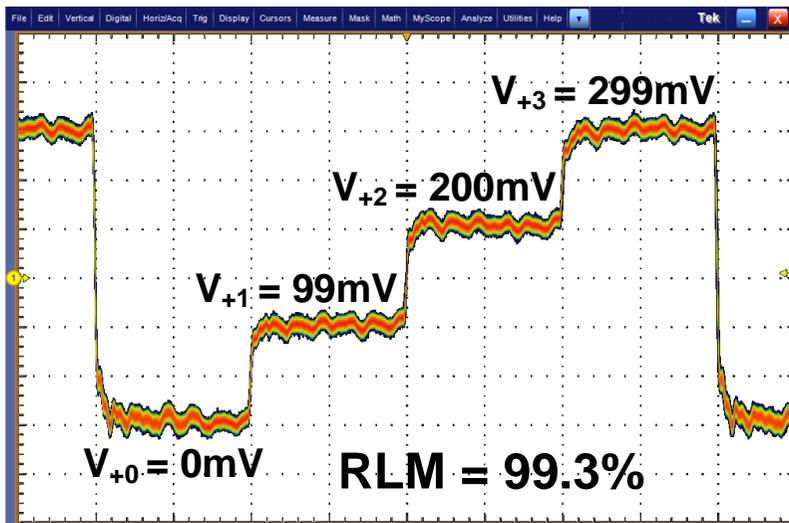


Figure 5.1.2. Measurement environment for the proposed PAM-4 transmitter.

Figure 5.1.2 shows the measurement setup for the proposed PAM-4 transmitter. Differential clock is generated from the clock source, MP1800A, and the PAM-4 output is measured using oscilloscope, MSO73304DX. Digital control codes are controlled using I2C.



(a)



(b)

Figure 5.1.3. Measured RLM with (a) one-point calibration and (b) three-point calibration.

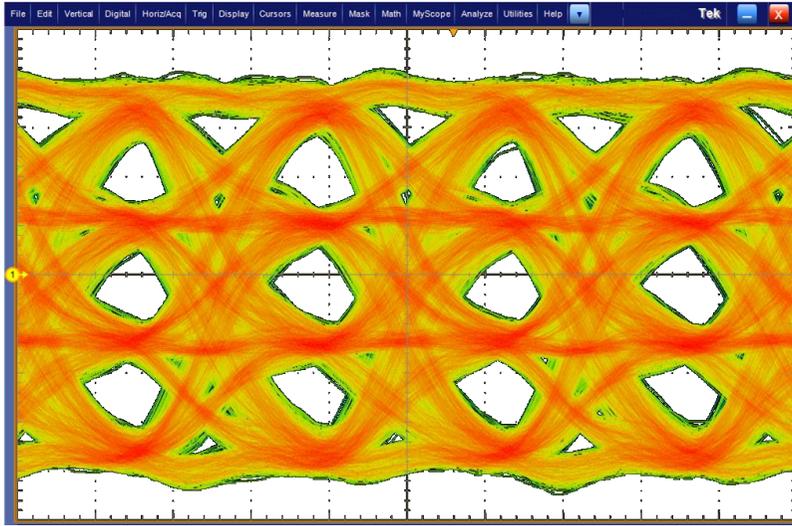


Figure 5.1.4. Measured eye diagram of the proposed PAM-4 transmitter at 28Gb/s with PRBS-7 pattern.

Figure 5.1.3 shows the measured RLMs of the proposed PAM-4 transmitter. With one-point calibration, which previous PAM-4 transmitters use, the measured RLM is 0.783. With the proposed 3-point calibration, the measured RLM is 0.993.

Figure 5.1.4 shows the measured eye diagram of the transmitter at 28Gb/s with PRBS-7 pattern. The measured eye has a width of 18.4ps and a height of 42.4mV. Figure 5.1.5 shows the measured bathtub curves at 28Gb/s for the PAM-4 output. The upper eye has BER of 10^{-12} with a timing margin of 0.18-UI. The middle eye has BER of 10^{-12} with a

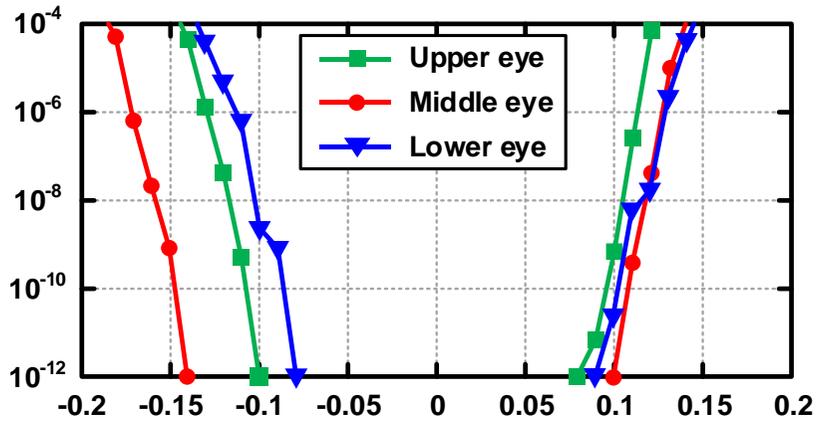


Figure 5.1.5. Bathtub curves for the proposed PAM-4 transmitter at 28Gb/s.

timing margin of 0.24-UI. The lower eye has BER of 10^{-12} with a timing margin of 0.17-UI. Therefore, the proposed transmitter has BER of 10^{-12} with a timing margin of 0.16-UI.

Figure 5.1.6 shows the power breakdown of the transmitter. At 28Gb/s, the transmitter consumes 17.89mW, and the driver consumes 1.78mW. The measured energy efficiency is 0.64pJ/bit.

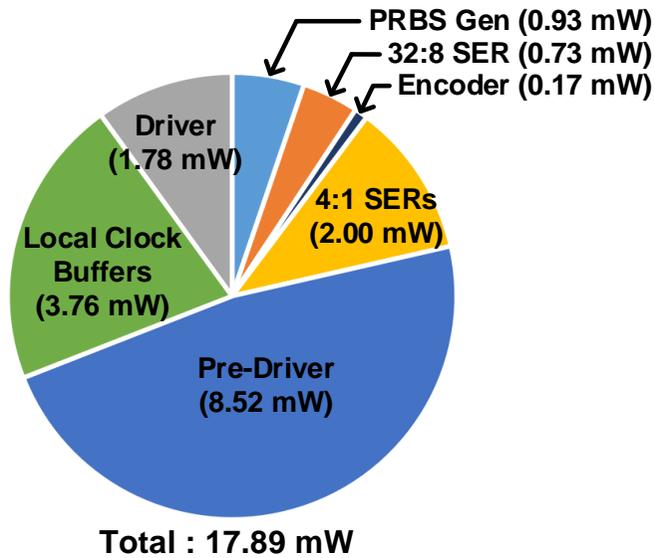


Figure 5.1.6. Power breakdown of the proposed PAM-4 transmitter at 28Gb/s.

Table 5.1.1 shows the performance summary of the proposed PAM-4 transmitter and comparison with other multi-level transmitters. By using a PAM-4, our transmitter has a pin efficiency of 200%. The RLM of 0.993 is achieved without using resistors and inductors. The proposed transmitter uniquely satisfies both accurate impedance matching and high linearity.

Table 5.1.1. Performance Summary and Comparison with Other Multi-Level Transmitters.

	This work	ISSCC 19'[2.1.2]	JSSC 14'[2.3.2]	ISSCC 19'[2.2.2.3]	JSSC 17'[3.3.1]
Technology	65nm	28nm	65nm	40nm	65nm
Data-rate per pin [Gb/s/pin]	28	27	7	56	16
Signaling	PAM-4	PAM-3	Duobinary	PAM-4	PAM-4
Pin efficiency	200%	150%	100%	200%	200%
Driver type	VM	VM (2-stcked)	CM	Diff. VM (w/ res. & inductor)	Diff. VM (w/ res. & inductor)
TX impedance matching	Yes	No	No	No	No
Multi-level linearity	Yes	Yes	Yes	No	No
PAM-4 impedance control	3-point ZQ calibration	.	.	1-point loop	1-point loop
PAM-4 RLM	0.993	.	.	0.977	0.967 ^a .
Energy efficiency	0.64	1.03 ^b .	0.09 ^c .	3.89	9.91
Area [mm ²]	0.033	0.0135 ^b .	0.0011 ^c .	0.56	0.0424

^a. Uses a look-up table

^b. Includes RX

^c. Pre-driver and driver only

5.2 REFLECTION-BASED DUOBINARY TRANSMITTER

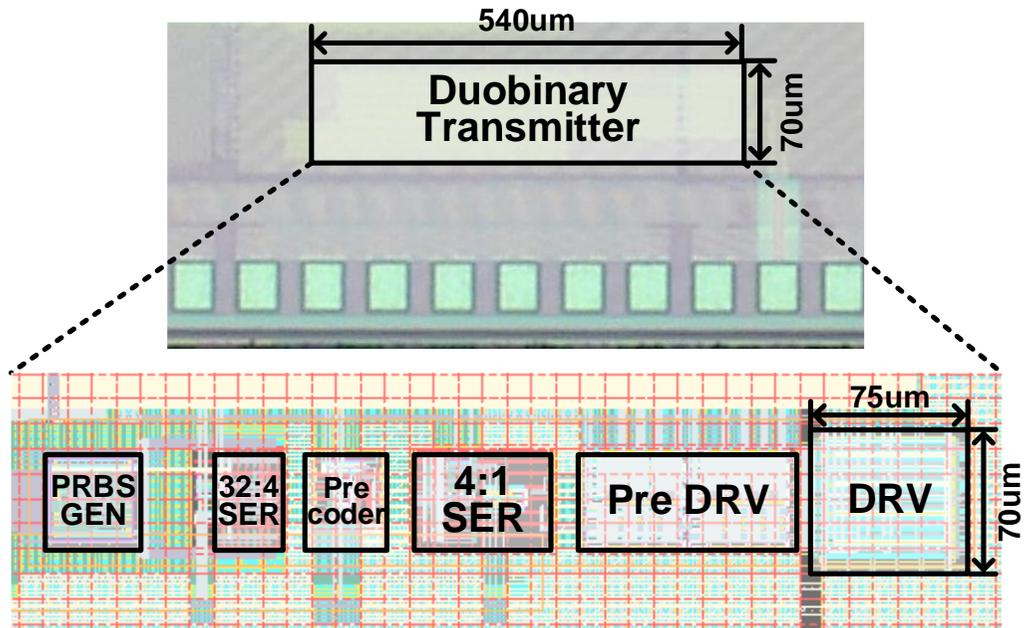


Figure 5.1.1. Chip micrograph of the proposed reflection-based duobinary transmitter.

The prototype chip of proposed reflection-based duobinary transmitter is fabricated in 65nm CMOS process. Figure 5.2.1 shows the micrograph and the total area of the proposed transmitter is 0.044mm^2 . Figure 5.2.2 shows the measurement setup for the proposed reflection-based duobinary transmitter. Differential clock is generated from the clock source, MP1800A, and the duobinary output is measured using oscilloscope, MSO73304DX. Digital control codes are controlled using I2C. The PCB board which has

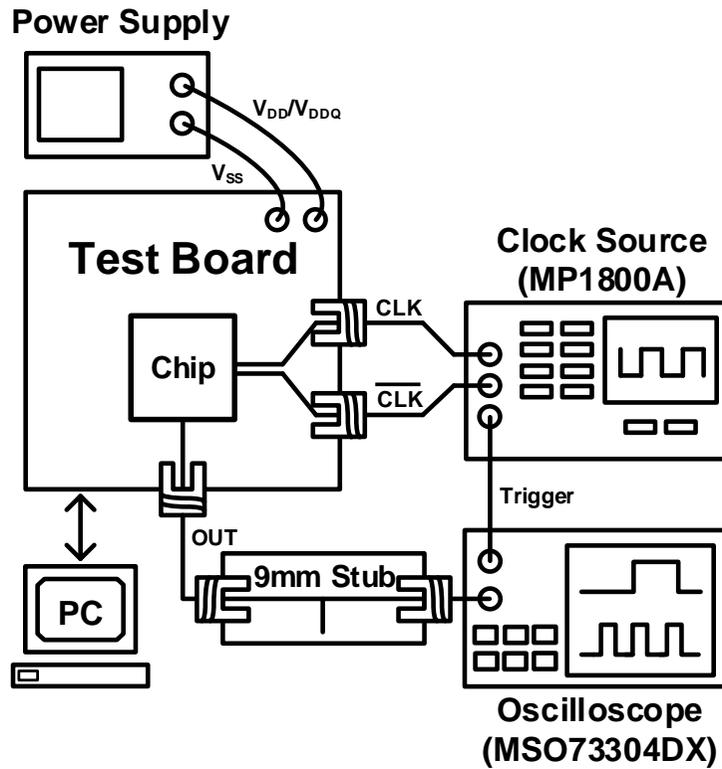
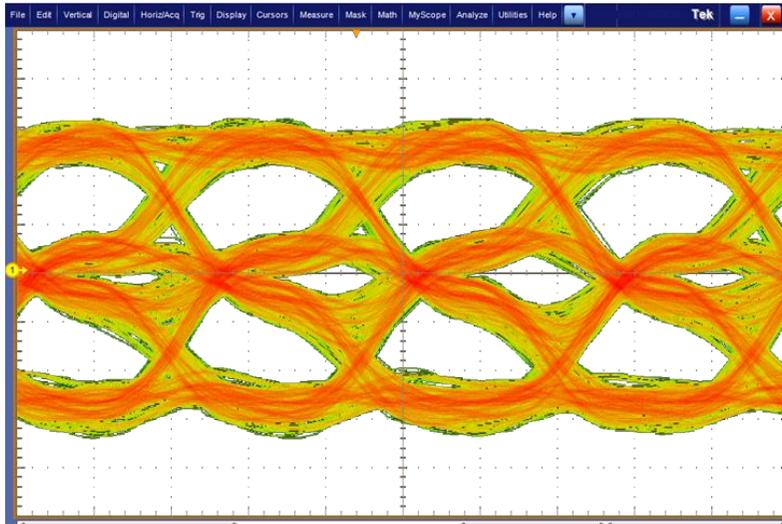


Figure 5.2.2. Measurement environment for the proposed reflection-based duobinary transmitter.

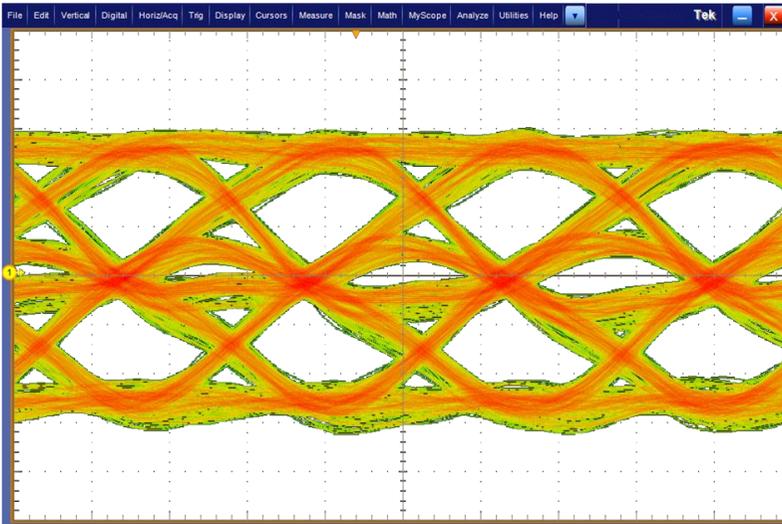
a 9mm stub is used for generating short reflection, and its flight time is about 55ps.

Figure 5.2.3, Figure 5.2.4, and Figure 5.2.5 show the measured eye diagrams with and without the equalization at 8Gb/s, 9Gb/s, and 10Gb/s, respectively. The eye diagrams are measured with the stub, which has flight time of about 60ps. To verify the proposed

transmitter the eye diagrams are measured in the three cases, $2 \cdot T_{\text{STUB}} < 1\text{-UI}$, $2 \cdot T_{\text{STUB}} \approx 1\text{-UI}$, and $2 \cdot T_{\text{STUB}} > 1\text{-UI}$. As shown in Figure 5.2.3, when $2 \cdot T_{\text{STUB}} < 1\text{-UI}$, the measured NRZ eye has a width of 26.0ps and a height of 13.2mV at 8Gb/s. The measured duobinary eye at 8Gb/s has a width of 85.0ps and a height of 86.4mV without the equalization and a width of 85.0ps and a height of 99.6mV with the equalization. As shown in Figure 5.2.4, when $2 \cdot T_{\text{STUB}} \approx 1\text{-UI}$, the measured NRZ eye has no eye opening at 9Gb/s. The measured duobinary eye at 9Gb/s has a width of 75.0ps and a height of 80.4mV without the equalization and a width of 77.5ps and a height of 85.2mV with the equalization. As shown in Figure 5.2.5, when $2 \cdot T_{\text{STUB}} > 1\text{-UI}$, the measured NRZ eye has no eye opening at 10Gb/s. The measured duobinary eye at 10Gb/s has a width of 50.8ps and a height of 66.0mV without the equalization and a width of 63.6ps and a height of 70.8mV with the equalization. At 10Gb/s, the transmitter consumes 13.8mW, and the driver consumes 1.8mW. The measured energy efficiency is 1.38pJ/bit.

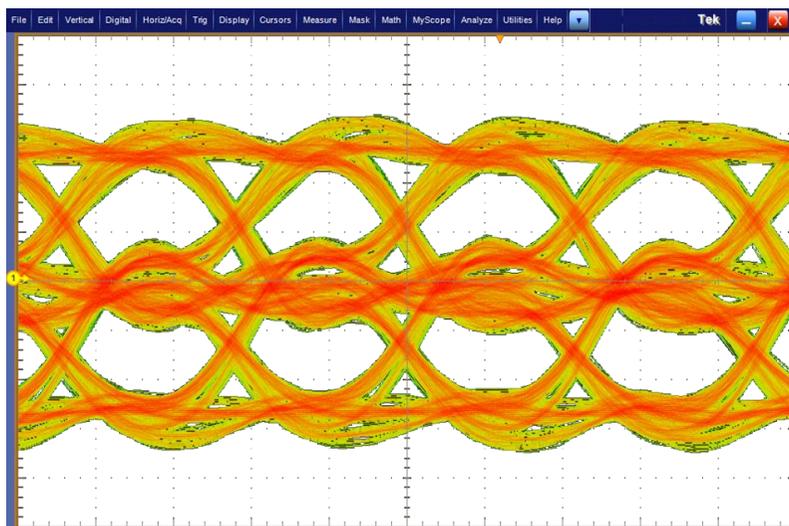


(a)

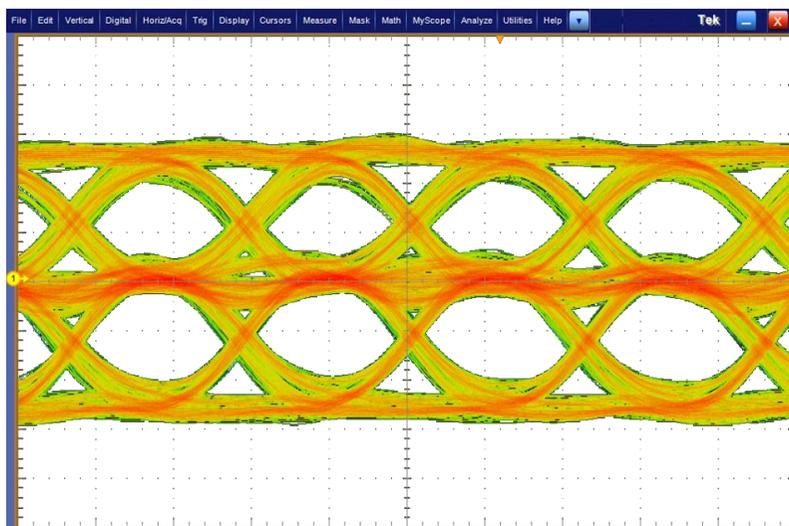


(b)

Figure 5.2.3. Measured eye diagram at 8Gb/s (a) without equalization and (b) with equalization.

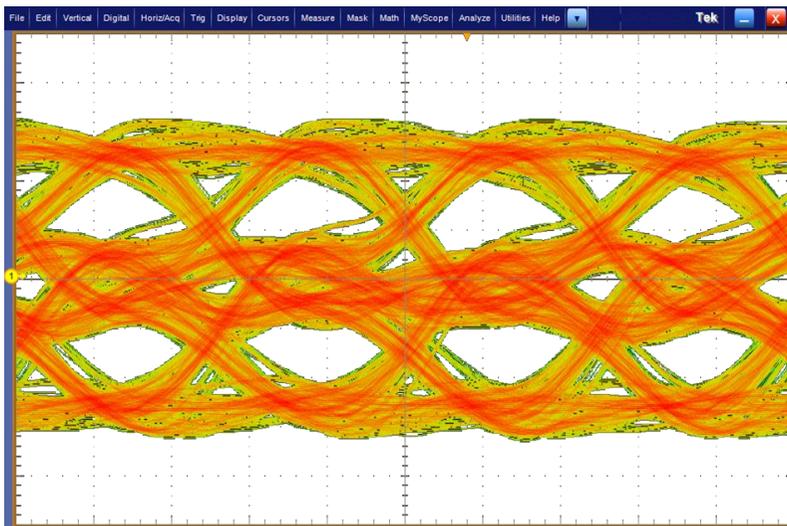


(a)

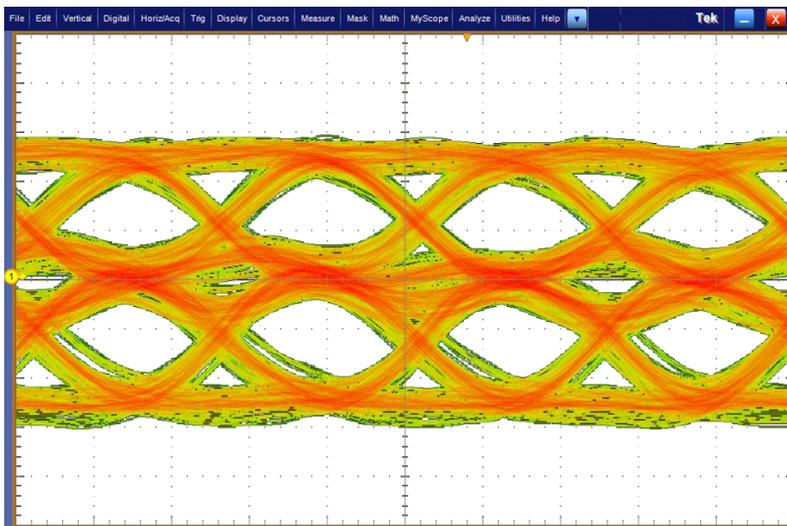


(b)

Figure 5.2.4. Measured eye diagram at 9Gb/s (a) without equalization and (b) with equalization.



(a)



(b)

Figure 5.2.5. Measured eye diagram at 10Gb/s (a) without equalization and (b) with equalization.

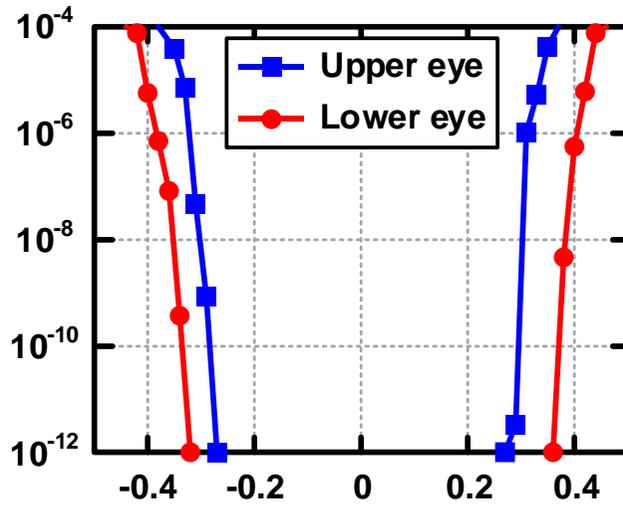


Figure 5.2.6. Bathtub curves for the proposed reflection-based duobinary transmitter at 10Gb/s.

Figure 5.2.6 shows the measured bathtub curves at 10Gb/s for the upper and lower eyes of the duobinary output. The upper eye has BER of 10^{-12} with a timing margin of 0.54-UI and the lower eye has BER of 10^{-12} with a timing margin of 0.65-UI. Therefore, the proposed transmitter has BER of 10^{-12} with a timing margin of 0.54-UI.

Table 5.2.1. Performance Summary and Comparison with Other Transmitters for Multi-Drop Systems.

	This work	JSSC 20'[4.2.1]	JSSC 15'[4.2.6]	ISSCC 11'[4.2.4]
Technology	65nm	1xnm DRAM	40nm	130nm
Supply voltage	1.0V/0.6V	1.05V	0.9V	1.2V
Target system	2-rank	2-rank	DIMM	DIMM
Signaling	Reflection-based duobinary	NRZ (w/ NT-ODT)	NRZ+QPSK	NRZ (w/ IMBM)
Data-rate per pin [Gb/s/pin]	10	7.5	3.75	2.4
Clock frequency [GHz]	2.5	1.875	5	2.4
Energy efficiency [pJ/b]	1.38	N/A	1.04	14.25
Area [mm ²]	0.0378	N/A	0.0051 ^a	0.336 ^b

^a. Active area

^b. Transceiver

Table 5.2.1 shows the performance summary of the proposed reflection-based duobinary transmitter and comparison with other transmitters for multi-drop systems. By using reflections, our transmitter achieves 10Gb/s in 2-rank memory system and energy efficiency of 1.38pJ/bit.

CHAPTER 6

CONCLUSION

In this thesis, two multi-level single-ended transmitters for memory interfaces have been proposed. The performance gap between processors and memory has been increased by 50% every year, making memory to be a bottle neck of the overall system. To increase memory bandwidth, the PAM-4 single-ended transmitter has been proposed. To compensate for the side effect of the multi-rank memory, the reflection-based duobinary transmitter has been proposed.

The proposed single-ended PAM-4 transmitter has the voltage-mode driver, which satisfies impedance matching and has high linearity. The driver has a resistorless and inductorless structure, thus occupying a small area. The proposed ZQ calibration for PAM-4 has three calibration points, which allow the transmitter to have accurate output impedance and linear output at all output levels. The ZQ calibration increases accuracy by considering impedance variation of both the driver and the receiver. A prototype has been fabricated in 65nm CMOS process, and the transmitter occupies 0.0333mm^2 . The measured eye has a width of 24.6ps and a height of 37.6mV at 28Gb/s, and the measured energy efficiency is 0.64pJ/b. The measured RLM with the 3-point ZQ calibration is 0.993.

To increase memory capacity, the stacked die packaging with multiple DRAM die stacked vertically in one package is widely used. However, combined with the center-pad structure, the structure creates stubs that cause short reflections and degrades signal integrity. We have proposed the reflection-based duobinary transmitter to mitigate this problem. The proposed transmitter utilizes reflection for duobinary signaling. The 2-tap opposite FFE and the slew-rate control are used to increase signal integrity. For testing, the PCB with short stub is used, and flight time of the stub is about 55ps. The measured duobinary eye at 8Gb/s has a width of 85.0ps and a height of 99.6mV while the NRZ eye has a width of 26ps and a height of 13.2mV. The measured duobinary eye at 9Gb/s has a width of 77.5ps and a height of 85.2mV while there is no NRZ eye opening. The measured duobinary eye at 10Gb/s has a width of 63.6ps and a height of 70.8mV while there is no NRZ eye opening. The measured energy efficiency is 1.38pJ/bit.

BIBLIOGRAPHY

- [1.1.1] K.-D. Hwang, B. Kim, S.-Y. Byeon, K.-Y. Kim, D.-H. Kwon, H.-B. Lee, G.-I. Lee, S.-S. Yoon, J.-Y. Cha, S.-Y. Jang, S.-H. Lee, Y.-S. Joo, G.-S. Lee, S.-S. Xi, S.-B. Lim, K.-H. Chu, J.-H. Cho, J. Chun, J. Oh, J. Kim, and S.-H. Lee, "A 16Gb/s/pin 8GB GDDR6 DRAM with bandwidth extension techniques for high-speed applications," in *IEEE International Solid-State Circuits Conference (ISSCC) Digest Technical Papers*, pp. 210-212, Feb. 2018.
- [1.1.2] C.-S. Oh, K. C. Chun, Y.-Y. Byun, Y.-K. Kim, S.-Y. Kim, Y. Ryu, J. Park, S. Kim, S. Cha, D. Shin, J. Lee, J.-P. Son, B.-K. Ho, S.-J. Cho, B. Kil, S. Ahn, B. Lim, Y. Park, K. Lee, M.-K. Lee, S. Kim, B. Lim, S.-K. Choi, J.-G. Kim, H.-I. Choi, H.-J. Kwon, J. J. Kong, K. Sohn, N. S. Kim, K.-I. Park, and J.-B. Lee, "A 1.1V 16Gb 640GB/s HBM2E DRAM with a data-bus window-extension technique and a synergetic on-die ECC scheme, " in *IEEE International Solid-State Circuits Conference (ISSCC) Digest Technical Papers*, pp. 330-332, Feb. 2020.
- [1.1.3] J. H. Cho, J. Kim, W. J. Lee, D. U. Lee, T. K. Kim, H. B. Park, C. Jeong, M.-J. Park, S. G. Baek, S. Choi, B. K. Yoon, Y. J. Choi, K. Y. Lee, D. Shim, J. Oh, J. Kim, and S.-H. Lee, "A 1.2V 64Gb 341GB/s HBM2 stacked DRAM with spiral point-to-point TSV structure and improved bank group data control," in *IEEE International Solid-State Circuits Conference (ISSCC) Digest Technical Papers*, pp. 208-210, Feb. 2018.
- [1.1.4] H. Kim, J. Cho, M. Kim, K. Kim, J. Lee, H. Lee, K. Park, K. Choi, H.-C. Bae, J. Kim, and J. Kim, "Measurement and Analysis of a High-Speed TSV Channel," in *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 2, no. 10, pp. 1672-1685, Oct. 2012.
- [1.1.5] C.-K. Lee, J. Lee, K. Kim, J.-S. Heo, J.-H. Baek, G.-H. Cha, D. Moon, D.-H. Lee, J.-W. Park, S. Lee, S.-H. Cho, Y.-R. Choi, K.-S. Ha, E. Seo, Y. Park, S.-J. Bae, I. Song, S.-H. Hyun, H.-J. Kwon, Y.-S. Sohn, J.-H. Choi, K.-I. Park, and S.-J. Jang,

- “Dual-loop two-step ZQ calibration for dynamic voltage-frequency scaling in LPDDR4 SDRAM,” *IEEE Journal of Solid-State Circuits*, vol. 53, no. 10, pp. 2906–2916, Oct. 2018.
- [2.1.1] J. Kim, A. Balankutty, R. K. Dokania, A. Elshazly, H. S. Kim, S. Kundu, D. Shi, S. Weaver, K. Yu, and F. O’Mahony, “A 112 Gb/s PAM-4 56 Gb/s NRZ reconfigurable transmitter with three-tap FFE in 10-nm FinFET,” *IEEE Journal of Solid-State Circuits*, vol. 54, no. 1, pp. 29–42, Jan. 2019.
- [2.1.2] H. Park, J. Song, Y. Lee, J. Sim, J. Choi, and C. Kim, “A 3-bit/2UI 27Gb/s PAM-3 single-ended transceiver using one-tap DFE for next generation memory interface,” in *IEEE International Solid-State Circuits Conference (ISSCC) Digest Technical Papers*, pp. 382-384, Feb. 2019.
- [2.1.3] B.-J. Yoo, D.-H. Lim, H. Pang, J.-H. Lee, S.-Y. Baek, N. Kim, D.-H. Choi, H. Yang, T. Yoon, D.-H. Choi, K. Kim, W. Jung, B.-K. Kim, J. Lee, G. Kang, S.-H. Park, M. Choi, and J. Shin, “A 56Gb/s 7.7mW/Gb/s PAM-4 wireline transceiver in 10nm FinFET using MM-CDR-Based ADC timing skew control and low-power DSP with approximate multiplier,” in *IEEE International Solid-State Circuits Conference (ISSCC) Digest Technical Papers*, pp. 122-124, Feb. 2020.
- [2.1.4] Aurangozeb, C. R. Dick, M. Mohammad, and M. Hossain, “Sequence-coded multilevel signaling for high-speed interface,” *IEEE Journal of Solid-State Circuits*, vol. 55, no. 1, pp. 27–37, Jan. 2020.
- [2.1.5] J. L. Zerbe, C. W. Werner, V. Stojanovic, F. Chen, J. Wei, G. Tsang, D. Kim, W. F. Stonecypher, A. Ho, T. P. Thrush, R. T. Kollipara, M. A. Horowitz, and K. S. Donnelly, “Equalization and clock recovery for a 2.5-10-Gb/s 2-PAM/4-PAM backplane transceiver cell,” *IEEE Journal of Solid-State Circuits*, vol. 38, no. 12, pp. 2121–2130, Dec. 2003.
- [2.1.6] J. Lee, P.-C. Chiang, P.-J. Peng, L.-Y. Chen, and C.-C. Weng, “Design of 56 Gb/s NRZ and PAM4 SerDes transceivers in CMOS technologies,” *IEEE Journal of Solid-State Circuits*, vol. 50, no. 9, pp. 2061–2073, Sep. 2015.
- [2.2.1.1] C. Hyun, H. Ko, J. Chae, H. Park and S. Kim, "A 20Gb/s dual-mode

- PAM4/NRZ single-ended transmitter with RLM compensation," *IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1-4, May 2019.
- [2.2.2.1] K.-L. J. Wong, H. Hatamkhani, M. Mansuri, and C.-K. K. Yang, "A 27-mW 3.6-Gb/s I/O transceiver," *IEEE Journal of Solid-State Circuits*, vol. 39, no. 4, pp. 602–612, Apr. 2004.
- [2.2.2.2] Y. Frans, J. Shin, L. Zhou, P. Upadhyaya, J. Im, V. Kireev, M. Elzeftawi, H. Hedayati, T. Pham, S. Asuncion, C. Borrelli, G. Zhang, H. Zhang, and K. Chang, "A 56-Gb/s PAM4 wireline transceiver using a 32-way time-interleaved SAR ADC in 16-nm FinFET," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 4, pp. 1101–1110, Apr. 2017.
- [2.2.2.3] P.-J. Peng, Y.-T. Chen, S.-T. Lai, C.-H. Chen, H.-E. Huang, and T. Shih, "A 112Gb/s PAM-4 voltage-mode transmitter with 4-tap two-step FFE and automatic phase alignment techniques in 40nm CMOS," in *IEEE International Solid-State Circuits Conference (ISSCC) Digest Technical Papers*, pp. 124-126, Feb. 2019.
- [2.2.3.1] M.-A. LaCroix, H. Wong, Y. H. Liu, H. Ho, S. Lebedev, P. Krotnev, D. A. Nicolescu, D. Petrov, C. Carvalho, S. Alie, E. Chong, F. A. Musa, and D. Tonietto, "A 60Gb/s PAM-4 ADC-DSP transceiver in 7nm CMOS with SNR-based adaptive power scaling achieving 6.9pJ/b at 32dB loss," in *IEEE International Solid-State Circuits Conference (ISSCC) Digest Technical Papers*, pp. 114-116, Feb. 2019.
- [2.3.1] H. Ko, J.-H. Chae, and S. Kim, "Single-ended voltage-mode duobinary transmitter with feedback time reduced parallel precoder," *Electronics Letters*, vol. 54, no. 15, pp. 936-937, Aug. 2018
- [2.3.2] S.-M. Lee, I.-M. Yi, H.-K. Jung, H. Lee, Y.-J. Kim, Y.-S. Kim, B. Kim, J.-Y. Sim, and H.-J. Park, "An 80 mV-swing single-ended duobinary transceiver with a TIA RX termination for the point-to-point DRAM interface," *IEEE Journal of Solid-State Circuits*, vol. 49, no. 11, pp. 2618–2630, Nov. 2014.
- [2.3.3] J. Lee, M. Chen and H. Wang, "Design and comparison of three 20-Gb/s

- backplane transceivers for duobinary, PAM4, and NRZ data," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 9, pp. 2120–2133, Sep. 2008.
- [3.3.1] A. Roshan-Zamir, O. Elhadidy, H.-W. Yang, and S. Palermo, "A reconfigurable 16/32 Gb/s dual-mode NRZ/PAM4 SerDes in 65-nm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 9, pp. 2430–2447, Sep. 2017.
- [4.1.1] P. J. Nair, D. A. Roberts and M. K. Qureshi, "Citadel: efficiently protecting stacked memory from large granularity failures," in *2014 47th Annual IEEE/ACM International Symposium on Microarchitecture*, pp. 51-62, Dec. 2014.
- [4.1.2] D. Kim, M. Park, S. Jang, J.-Y. Song, H. Chi, G. Choi, S. Choi, C. Kim, M. Han, K. Koo, Y. Kim, D. U. Lee, J. Lee, K. Kwon, B. Choi, H. Kim, S. Ku, J. Kim, S. Oh, D. Im, Y. Lee, M. Park, J. Choi, J. Chun, and K. Jin, "A 1.1-V 10-nm class 6.4-Gb/s/pin 16-Gb DDR5 SDRAM with a phase rotator-ILO DLL, high-speed SerDes, and DFE/FFE equalization scheme for Rx/Tx," *IEEE Journal of Solid-State Circuits*, vol. 55, no. 1, pp. 167–177, Jan. 2020.
- [4.1.3] J. Seo, J. Ko, S. Lee, J.-Y. Sim, H.-J. Park, and B. Kim, "A 7.8-Gb/s 2.9-pJ/b single-ended receiver with 20-Tap DFE for highly reflective channels," in *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 28, no. 3, pp. 818-822, Mar. 2020.
- [4.2.1] K.-S. Ha, C.-K. Lee, D. Lee, D. Moon, H.-R. Hwang, D. Park, Y.-H. Kim, Y. H. Son, B. Na, S. Lee, Y.-S. Park, H.-J. Kwon, T.-Y. Oh, Y.-S. Sohn, S.-J. Bae, K.-I. Park, and J.-B. Lee, "A 7.5 Gb/s/pin 8-Gb LPDDR5 SDRAM with various high-speed and low-power techniques," *IEEE Journal of Solid-State Circuits*, vol. 55, no. 1, pp. 157–166, Jan. 2020.
- [4.2.2] K.-S. Ha, C.-K. Lee, D. Lee, D. Moon, J.-H. Jang, H.-R. H, H. Chi, J. Park, Y. Choi, Y.-H. Kim, Y. Son, H. Cho, B. Na, H.-J. Ahn, S. Lee, S.-K. Choi, Y.-S. Park, S.-H. Hyun, S. Chang, H.-J. Kwon, J.-H. Choi, T.-Y. Oh, Y.-S. Sohn, K.-I. Park, and S.-J. Jang, "A 7.6Gb/s/pin LPDDR5 SDRAM with WCK clocking and non-target ODT for high speed and with DVFS, internal data copy, and deep-sleep mode for low power," in *IEEE International Solid-State Circuits Conference*

(ISSCC) Digest Technical Papers, pp. 378-380, Feb. 2019.

- [4.2.3] S. Gupta, "Non-target DRAM termination in high speed LPDDR system for improved signal integrity," in *IEEE Symposium on Electromagnetic Compatibility, Signal Integrity and Power Integrity (EMC, SI & PI)*, Jul. 2018.
- [4.2.4] W.-Y. Shin, G.-M. Hong, H. Lee, J.-D. Han, S. Kim, K.-S. Park, D.-H. Lim, J.-H. Chun, D.-K. Jeong, and S. Kim, "A 4.8Gb/s impedance-matched bidirectional multi-drop transceiver for high-capacity memory interface," in *IEEE International Solid-State Circuits Conference (ISSCC) Digest Technical Papers*, pp. 494-496, Feb. 2011.
- [4.2.5] Y. Yoon, and D.-K. Jeong, "A multidrop bus design scheme with resistor-based impedance matching on nonuniform impedance lines," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 58, no. 6, pp. 1264-1276, Jun. 2011.
- [4.2.6] K. Gharibdoust, A. Tajalli, and Y. Leblebici, "Hybrid NRZ/multi-tone serial data transceiver for multi-drop memory interfaces," *IEEE Journal of Solid-State Circuits*, vol. 50, no. 12, pp. 3133-3144, Dec. 2015.
- [4.3.3.1] J. Park, J.-H. Chae, Y.-U. Jeong, J.-W. Lee, and S. Kim, "A 2.1-Gb/s 12-channel transmitter with phase emphasis embedded serializer for 55-in UHD intra-panel interface," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 10, pp. 2878-2888, Oct. 2018.
- [4.3.3.2] J.-H. Chae, M. Kim, G.-M. Hong, J. Park, and S. Kim, "A 3.2 Gb/s 16-channel transmitter for intra-panel interfaces, with independently controllable output swing, common-mode voltage, and equalization," *IEEE Access*, vol. 6, no. 1, pp. 78055-78064, Dec. 2018.

한글초록

본 연구에서 메모리 인터페이스를 위한 멀티 레벨 송신기가 제시되었다. 프로세서와 메모리 간의 성능 차이가 매년 계속 증가함에 따라, 메모리는 전체 시스템의 병목점이 되고있다. 우리는 메모리 대역폭을 늘리기 위해 PAM-4 단일 종단 송신기를 제안하였고, 멀티 랭크 메모리를 위한 duobinary 단일 종단 송신기를 제안하였다.

제안된 PAM-4 송신기의 드라이버는 높은 선형성과 임피던스 정합을 동시에 만족한다. 또한 저항이나 인덕터를 사용하지 않아 작은 면적을 차지한다. 제안된 ZQ 캘리브레이션은 세개의 교정 점을 가지고 있어 송신기가 정확한 임피던스와 선형적인 출력을 갖게 한다. 프로토 타입은 65nm CMOS 공정으로 제작되었고 송신기는 0.0333mm^2 의 면적을 차지한다. 측정된 28Gb/s에서의 eye는 18.3ps의 길이와 42.4mV의 높이를 갖고, 에너지 효율은 0.64pJ/bit이다. ZQ 캘리브레이션과 함께 측정된 RLM은 0.993이다.

메모리의 용량을 늘리기 위해 하나의 패키지에 여러 개의 DRAM 다이를 수직으로 쌓는 패키징은 메모리의 중앙 패드 구조와 결합되어 짧은 반사를 야기하는 스텝을 만든다. 우리는 이 문제를 완화하기 위해 반사 기반 duobinary 송신기를 제안했다. 이 송신기는 반사를 이용하여 duobinary signaling을 한다. 2 탭 반대 강조 기술과 슬루 레이트 조절 기술이 신호 완결성을 높이기 위해 사용되었다. NRZ eye가 없는 10Gb/s에서 측정된 duobinary eye는 63.6ps 길이와

70.8mV의 높이를 갖는다. 측정된 에너지 효율은 1.38pJ/bit이다.

주요어 : 메모리 인터페이스, PAM-4 송신기, ZQ 캘리브레이션, duobinary 송신기,
출력 드라이버

학 번 : 2013-20879