



Ph.D.Dissertation

Design of Low-Power Transceiver for Memory Interface

메모리 인터페이스 용

저전력 송수신기 설계

by

Jung-Hun Park

February, 2023

Department of Electrical and Computer Engineering College of Engineering Seoul National University

Design of Low-Power Transceiver for Memory Interface

지도 교수 정 덕 균

이 논문을 공학박사 학위논문으로 제출함 2023 년 2 월

> 서울대학교 대학원 전기·정보공학부 박 정 훈

박정훈의 박사 학위논문을 인준함 2023 년 2 월



Design of Low-Power Transceiver for Memory Interface

by

Jung-Hun Park

A Dissertation Submitted to the Department of Electrical and Computer Engineering in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

at

SEOUL NATIONAL UNIVERSITY

February, 2023

Committee in Charge:

Professor Jaeha Kim, Chairman

Professor Deog-Kyoon Jeong, Vice-Chairman

Professor Woo-Seok Choi

Professor Woogeun Rhee

Professor Yongsam Moon

Abstract

This thesis presents design techniques for low-power transceivers for memory interfaces. In terms of two trends to improve the bandwidth of the memory interface, fast-and-narrow and wide-and-slow, methods for minimizing power consumption have been studied.

First, methods for optimizing the power consumption of the HBM interface are studied. A training sequence is introduced to efficiently optimize a large number of transceivers. The strengths of the drivers are calibrated and the reference voltages of the samplers are adjusted through DC-based training. SBR-based training enables DQS alignment and FFE coefficient optimization in a much shorter time than 2-D eye monitoring methods. Through the training sequence, 8 PAM-4 transceivers are optimized within 1-ms and satisfies the BER < 10⁻¹² even at low VDDQ. In addition, the proposed charge-recycling latch saves power consumption of samplers by 44.5% and enables high-speed operation by reducing decision time. With the help of the training sequence and the charge-recycling latch, the proposed HBM interface achieved 68.7-fJ/b/mm, which is the best energy efficiency comparing to that of state-of-the-art memory interfaces, and the second best performance to that of recently published on-chip serial links.

Second, methods of minimizing the area and power consumption of transmitters for high-bandwidth-per-pin memory interfaces are studied. The proposed PN-over-NP driver enables 50Ω matching without series resistors, reducing the area of the driver and saving the power consumption of the driver and pre-driver. In addition, the T-coil-combined edge-boosting equalizer eliminates unnecessary current waste of the FFE to minimize power consumption when there is no transition, while maintaining output impedance at high frequencies to improve signal integrity. In addition, a CMOS-based clock error corrector that does not use passive elements is used to effectively calibrate a 4-phase clock using only a small area. Thanks to the proposed structures of the driver and the equalizer, the proposed transmitter achieves a power efficiency of 0.51pJ/b, which is the best compared to state-of-the-art single-ended transmitters including an equalizer. Area of the transmitter is 5008um² including T-coil.

Keywords : HBM, GDDR, memory interface, on-chip training, N-over-N driver, feed-forward equalizer (FFE), DQS alignment, offset calibration, single-ended signaling, impedance matching, T-coil, edge-boosting equalizer.

Student Number : 2018-27402

Contents

ABSTRACT	Ι
CONTENTS	III
LIST OF FIGURES	V
LIST OF TABLES	VIII
CHAPTER 1 INTRODUCTION	1
1.1 MOTIVATION	1
1.2 THESIS ORGANIZATION	6
CHAPTER 2 BACKGROUND ON LOW-POWER HIGH-	SPEED MEMORY
INTERFACES	8
2.1 Overview	8
2.2 FAST AND NARROW MEMORY INTERFACE	11
2.3 WIDE AND SLOW MEMORY INTERFACE	14
CHAPTER 3 BACKGROUND ON VOLTAGE-MODE DR	IVERS FOR
MEMORY INTERFACES	20
3.1 Overview	20
3.2 OUTPUT IMPEDANCE	23
CHAPTER 4 HBM INTERFACE WITH PER-PIN SELF-7	FRAINING
SEQUENCE	26

5.4.1 OVERALL ARCHITECTURE 5.4.2 CMOS-BASED CLOCK ERROR CORRECTOR 5.5 MEASUREMENT RESULT CHAPTER 6 CONCLUSION BIBLIOGRAPHY	72 75 79 87 89
 5.4.1 OVERALL ARCHITECTURE 5.4.2 CMOS-BASED CLOCK ERROR CORRECTOR 5.5 MEASUREMENT RESULT CHAPTER 6 CONCLUSION 	72
 5.4.1 OVERALL ARCHITECTURE 5.4.2 CMOS-BASED CLOCK ERROR CORRECTOR 5.5 MEASUREMENT RESULT 	72
5.4.1 OVERALL ARCHITECTURE 5.4.2 CMOS-based Clock Error Corrector	
5.4.1 OVERALL ARCHITECTURE	12
	70
5.4 CIRCUIT IMPLEMENTATION	72
5.3 PROPOSED T-COIL-COMBINED EDGE-BOOSTING EQUALIZER	66
5.2 PROPOSED PN-NP DRIVER	60
5.1 Overview	58
T-COIL-COMBINED EDGE-BOOSTING EQUALIZER	58
CHAPTER 5 LOW-POWER TRANSMITTER WITH PN-NP DRIV	ER AND
4.4 Measurement Result	49
4.3.3 SBR-BASED TRAINING	47
4.3.2 DC-BASED TRAINING	40
4.3.1 OVERALL TRAINING SEQUENCE	36
4.3 PROPOSED PER-PIN TRAINING SEQUENCE	36
4.2.2 CHARGE-RECYCLING LATCH	
4.2.1 LOW-POWER SINGLE-ENDED PAM-4 TRANSCEIVER	
4.2 PROPOSED HBM INTERFACE	

List of Figures

FIG. 1.1 STANDARDS OF DDR, LPDDR, GDDR, AND HBM2
FIG. 1.2 COMPARISON OF CONVENTIONAL APPROACH VERSUS IMC [3]
FIG. 1.3 ARCHITECTURE OF FIMDRAM BASED ON HBM2 [4]
FIG.1.4 ASYMMETRIC T-COIL FOR GDDR6 I/O [9]5
FIG. 1.5 FFE-COMBINED XTC SCHEME [10]
FIG. 2.1 THE DRAM DATA BANDWIDTH GROWTH [11]9
FIG. 2.2 THE DRAM PER-PIN BANDWIDTH HISTORY [12]9
FIG. 2.3 WCK CLOCKING OF GDDR6 [60]12
FIG. 2.4 TRAINING SEQUENCE OF GDDR6 [59]13
FIG. 2.5 SCHEMATIC OF THE EMIB [62]14
FIG. 2.6 Illustration of eight-high stacked HBM in a 2.5-D SiP [56]17
FIG. 2.7 HIGH LEVEL BLOCK DIAGRAM EXAMPLE OF CLOCKING SCHEME [85]18
FIG. 2.8 INITIALIZATION SEQUENCE WITH LANE REPAIRS [85]19
FIG. 3.1 THE PULL-UP AND PULL-DOWN OPERATION OF (A) SSTL AND (B) POD21
FIG. 3.2 THE PULL-UP AND PULL-DOWN OPERATION OF (A) HSUL AND (B) LVSTL22
FIG. 3.3 Two methods for impedance calibration [104]23
FIG. 3.4 OUTPUT IMPEDANCE OF N-OVER-N DRIVER
FIG. 3.5 IMPEDANCE-MATCHED PAM-4 N-OVER-N DRIVER AND ITS OPERATION [105]25
FIG. 4.1 OVERALL ARCHITECTURE OF THE PROPOSED HBM INTERFACE
Fig. 4.2 Schematic of the 3-b phase selector and 6-b phase interpolator
FIG. 4.3 POST-LAYOUT SIMULATION RESULT OF THE PHASE INTERPOLATOR

FIG. 4.4 SCHEMATIC OF THE PAM-4 TRANSMITTER	32
FIG. 4.5 SCHEMATIC OF CONVENTIONAL PMOS STRONGARM LATCH	33
FIG. 4.6 SCHEMATIC OF THE PROPOSED CHARGE-RECYCLING LATCH	34
FIG. 4.7 VOLTAGES OF OUTPUT NODES DURING RESET AND DECISION PHASE	35
Fig. 4.8 Comparison of full-scanning, $1x2y3x$ iteration, and SBR-based method	37
FIG. 4.9 CONCEPTUAL DIAGRAM OF THE ENTIRE TRAINING SEQUENCE	38
FIG. 4.10 DQS LANES AND THE SIDEBAND FOR TRAINING	39
FIG. 4.11 CONCEPTUAL DIAGRAM OF THE DC-BASED RLM TRAINING	40
FIG. 4.12 PAM-4 N-OVER-N DRIVER WITH 2-TAP FFE AND PAM-4 DC LEVEL	41
FIG. 4.13 EYE HEIGHT VERSUS PULL-UP TO PULL-DOWN RATIO	43
FIG. 4.14 SIMULATION RESULT OF THE RLM TRAINING	43
FIG. 4.15 TRAINING BLOCKS USED FOR IMPEDANCE CALIBRATION	44
FIG. 4.16 MONTE-CARLO SIMULATION RESULT OF THE LOW-POWER SAMPLER	45
FIG. 4.17 TRAINING BLOCKS FOR OFFSET CALIBRATION	46
FIG. 4.18 CONCEPTUAL DIAGRAM OF OFFSET CALIBRATION	46
FIG. 4.19 TRAINING BLOCKS FOR EYE CENTERING	47
FIG. 4.20 SINGLE-BIT RESPONSE BEFORE AND AFTER FFE COEFFICIENT TRAINING	48
FIG. 4.21 CHIP PHOTOMICROGRAPH	50
FIG. 4.22 CROSS SECTION VIEW AND INSERTION LOSS OF THE METAL CHANNEL	51
FIG. 4.23 MEASUREMENT SETUP	51
FIG. 4.24 EYE DIAGRAMS BEFORE AND AFTER TRAINING AT VDDQ=0.6V	53
FIG. 4.25 BER BATHTUB AT VDDQ=0.6V	53
FIG. 4.26 EYE DIAGRAMS BEFORE AND AFTER TRAINING AT VDDQ=0.55V	54
FIG. 4.27 BER BATHTUB AT VDDQ=0.55V	54

FIG. 4.28 AREA AND POWER BREAKDOWN	55
FIG. 5.1 PROPOSED PN-OVER-NP DRIVER OPERATION	61
FIG. 5.2 OUTPUT IMPEDANCE OF THE PN-OVER-NP DRIVER	63
FIG. 5.3 SWING CHARACTERISTICS OF THE PROPOSED PN-OVER-NP DRIVER	65
FIG. 5.4 IMPLEMENTATION OF THE T-COIL	67
FIG. 5.5 SCHEMATIC OF SYMMETRIC T-COIL-BASED EDGE-BOOSTING EQUALIZER	68
FIG. 5.6 RLC EQUIVALENT MODEL OF THE PROPOSED EQUALIZER	68
FIG. 5.7 OUTPUT IMPEDANCE OF THE TRANSMITTER	69
FIG. 5.8 RLC EQUIVALENT MODEL OF THREE SITUATIONS	71
FIG. 5.9 OUTPUT IMPEDANCE AND GAIN DEPENDING ON THE FREQUENCY	71
FIG. 5.10 OVERALL ARCHITECTURE OF THE PROPOSED TRANSMITTER	73
FIG. 5.11 HIGH-SPEED 4:1 SERIALIZER (MUX)	74
FIG. 5.12 CMOS-BASED CLOCK ERROR CORRECTOR	76
FIG. 5.13 POST-LAYOUT SIMULATION RESULT OF THE CLOCK ERROR CORRECTOR	78
FIG. 5.14 EYE DIAGRAM (A) WITH AND (B) WITHOUT CLOCK ERROR CORRECTOR	78
FIG. 5.15 CHIP PHOTOMICROGRAPH AND AREA BREAKDOWN	80
FIG. 5.16 MEASUREMENT SETUP	82
FIG. 5.17 CHARACTERISTICS OF THE EXTERNAL CLOCK	83
FIG. 5.18 S PARAMETERS OF THE 8MM FR-4 TRACE	83
Fig. 5.19 Eye diagrams measured at 12.8, 20.0, and 32.0 Gb/s	84
FIG. 5.20 POWER BREAKDOWN	85

VIII

List of Tables

TABLE 2.1 THE FEATURE COMPARISON – GDDR5 VS GDDR5X VS GDDR6 [60]	11
TABLE 2.2 SUMMARY OF HBM1~HBM3 [52]-[58]	16
TABLE 4.1 COMPARISON WITH MEMORY INTERFACES	57
TABLE 4.2 COMPARISON WITH ON-CHIP SERIAL LINKS	57
TABLE 5.1 OPERATION PRINCIPLE OF THE ERROR CORRECTOR	77
TABLE 5.2 PERFORMANCE COMPARISON TABLE FOR STATE-OF-THE-ART SINGLE-ENDED TX-	
EQUALIZED TRANSMITTERS	86

Chapter 1

Introduction

1.1 Motivation

In 1966, IBM's Robert H. Dennard proposed a 1T1C memory cell that can store digital data using a single transistor and a single capacitor instead of SRAM, which stores data using six transistors, and this structure became the origin of today's DRAM [1]. The simple structure of DRAM enables high integration, good power efficiency, and low price, thus it is used as most of the main memory devices today. Since 2000 when the double-data rate synchronous DRAM (DDR SDRAM) standard was announced, DDR has been developed in a very diverse domains. Fig. 1.1 summarizes the years in which standards of DDR, Low-Power DDR (LPDDR), Graphics DDR (GDDR), and High-Bandwidth Memory (HBM) were announced.

YEAR	DDR	LPDDR	GDDR	HBM
2000	DDR			
2001				
2002				
2003	DDR2		GDDR3	
2004				
2005			GDDR4	
2006		LPDDR		
2007	DDR3			
2008			GDDR5	
2009		LPDDR2		
2010	DDR3U			
2011	DDR3L	LPDDR3		
2012	DDR4			
2013		LPDDR4		HBM
2014				
2015				
2016		LPDDR4X	GDDR5X	HBM2
2017				
2018			GDDR6	HBM2E
2019		LPDDR5		
2020	DDR5		GDDR6X	
2021		LPDDR5X		
2022				HBM3

Fig. 1.1 Standards	s of DDR, LPDDR,	GDDR, and HBM
--------------------	------------------	---------------

Despite the rapid increase in DRAM bandwidth, the amount of required data has increased at a faster rate due to the development of big data and artificial intelligence (AI) technology. Therefore, the DRAM interface has become a bottleneck in high-speed systems. In 1995, Wulf and Mckee used the term "Memory Wall" to present problems caused by the speed of DRAM not catching up with the processor [2] . This problem remains unresolved to this day, and attempts have emerged to perform operations in memory cells in order to minimize the burden on the memory interface. In-memory computing (IMC), which performs simple matrix calculations within SRAM [3] and function-in-memory DRAM (FIM-DRAM), which places a computing unit inside the DRAM stack of HBM [4] , have been proposed as alternatives to the von Neumann architecture. However, these attempts have not yet been commercialized due to problems such as memory capacity, heat generation, and price, and their applications are limited to specific domains.



Fig. 1.2 Comparison of conventional approach versus IMC [3]

Therefore, efforts are being made to improve the bandwidth of DRAM. A typical method is to apply multi-level signaling. Since the DRAM process has lower performance than the logic process, operating frequency is limited. Therefore, methods such as duobinary [5], QAM-16 [6], PAM-3 [7], and PAM-4 [8] that can increase data bandwidth without changing the Nyquist frequency have been proposed for low-power memory interfaces. In addition, a method of designing a T-coil using the RDL layer of GDDR was proposed, which could mitigate the bandwidth reduction due to the capacitance of the ESD protection circuit and the receiver [9]. Methods to increase data throughput instead of bandwidth per pin have also been proposed. For example, for HBM interface, a method of effectively transmitting data at a narrow channel pitch by removing crosstalk using FFE-combined crosstalk cancellation scheme (XTC) was proposed [10].

Not only bandwidth but also power efficiency is an important issue of DRAM interface. In general, transceiver bandwidth is in a trade-off relationship with power consumption, so an increase in DRAM bandwidth causes an increase in power con-



Fig. 1.3 Architecture of FIMDRAM based on HBM2 [4]

sumption and heat generation of the entire system. For instance, since the DRAM interface uses a large number of DQ pins, high transceiver power consumption deteriorates signal integrity due to simultaneous switching output (SSO) noise. In addition, the heat of the interface circuit deteriorates the performance of the DRAM cell. Therefore, a method to minimize the power of the transceiver is required. In this thesis, a method for improving bandwidth while reducing power consumption of the DRAM interface and wide-and-slow interface.



Fig.1.4 Asymmetric T-coil for GDDR6 I/O [9]



Fig. 1.5 FFE-combined XTC scheme [10]

1.2 Thesis Organization

This thesis is organized as follows. Chapter 2 describes the backgrounds of low-power high-speed memory interfaces. The recent data bandwidth trend of DRAM interface and two approaches for high bandwidth, fast-and-narrow memory interface and wide-and-slow memory interface are briefly introduced.

The features of the voltage-mode drivers for memory interfaces are described in Chapter 3. As the memory interfaces are developed and subdivided, various types of voltage mode drivers and termination methods suitable for the memory applications have been introduced. In this chapter, the characteristics of each driver and termination methodologies are described. It is also briefly explained how the output impedance of the P-over-N driver and the N-over-N driver are calibrated for impedance matching.

In Chapter 4, the proposed HBM interface and per-pin training sequence are described. First, the reasons why low-power is important for HBM interface and why a per-pin training sequence is necessary to achieve low-power are briefly explained. Next, the circuit elements of the proposed HBM interface and the structure and operation principle of the charge-recycling latch are explained. Finally, a per-pin training sequence for optimizing the interface is described. Through the comparison of transceiver performance before and after training, the effectiveness of the proposed training sequence is verified, and the performance of the transceiver is compared with state-of-the-art memory interfaces and on-chip serial links.

Chapter 5 introduces a low-power transmitter with high bandwidth per pin. First, the structures and operating principles of the proposed PN-over-NP driver and T-coil-combined edge-boosting equalizer are explained. Then, the CMOSbased clock error corrector and the structure of the entire transmitter are explained. The proposed driver and equalizer are verified by comparing the performance of the transmitter with state-of-the-art single-ended transmitters including equalizers.

Chapter 6 summarizes the proposed works and concludes this thesis.

Chapter 2

Background on Low-Power High-Speed Memory Interfaces

2.1 Overview

Fig. 2.1 shows the DRAM data bandwidth growth by year [11]. The bandwidth of DRAM has increased on a log scale to respond to market demand. Therefore, as shown in Fig. 2.2, the per-pin bandwidth of DRAM also increased rapidly [12]. In the early 2000s, the per-pin bandwidth of DDR was around 1Gb/s/pin, but has increased to 6.4Gb/s/pin [13] -[19]. Bandwidth per pin of LPDDR3 was also similar to that of DDR at 1.6Gb/s/pin, but in 2022, LPDDR5X with a per-pin bandwidth of 9.5Gb/s/pin was presented to the academic world [20] -[31]. GDDR, which requires the highest data bandwidth, had the same speed as DDR in 2000,







Fig. 2.2 The DRAM per-pin bandwidth history [12]

but since then it has risen rapidly [32] -[46]. In 2022, a GDDR6 interface with a per-pin bandwidth of 27Gb/s/pin was presented [9].

However, the difference in performance between the DRAM process and the logic process causes limitations of this approach. In contrast to recent ultra-high-speed wireline whose bandwidth per pin have risen to 100+Gb/s/pin, GDDR's bandwidth per pin is lower than 30Gb/s/pin. A 60-Gb/s/pin PAM-4 transmitter [47] and a 40-Gb/s/pin PAM-4 transceiver [48] for memory interfaces have been presented in the academia, but they were fabricated in a mimicked 10-nm DRAM process using a 28-nm CMOS process. To overcome these limitations, a wide-and-slow approaches have been proposed. Wide-IO DRAMs using through silicon via (TSV) have been proposed [49] -[51], and HBM has improved its bandwidth to hundreds of GB/s with high data throughput despite of its low per-pin bandwidth [4], [52] -[58].

In this chapter, the two approaches to improve DRAM bandwidth and their key characteristics are described.

2.2 Fast and Narrow Memory Interface

Fast-and-narrow memory interfaces have evolved in the direction of increasing bandwidth per pin, like general high-speed wirelines. DDR, GDDR that requires the highest performance, and LPDDR that consumes low power could be classified into a fast-and-narrow approach. Table 2.1 shows the key features of GDDR5, GDDR5X, and GDDR6. According to the JEDEC standard, per-pin bandwidth of GDDR6 is 16Gb/s/pin, which is close to twice that of LPDDR5X and GDDR5 [59], [60].

GDDR transmits data using a high VDD of 1.25V or 1.35V. Therefore, various techniques used in high speed wirelines could be applied. The most representative is pulse-amplitude modulation. Since PAM-4 signaling has twice the data rate

Feature	GDDR5	GDDR5X	GDDR6
Density	(2Gb) 4Gb, 8Gb	8Gb	8Gb, 16Gb (32Gb)
VDD, VDDQ	1.5V, 1.35V	1.35V	1.35V, 1.25V
Vpp	N/A	1.8V	1.8V
Data rates	Up to 8Gb/s	Up to 12Gb/s	Up to 16Gb/s
Channel count	1	1	2
Burst length	8	16/8	16
I/O width	x32/x16	x32/x16	2ch x 16/x8
DFE	N/A	N/A	1-tap DFE

Table 2.1 The feature comparison – GDDR5 vs GDDR5X vs GDDR6 [60]

at the same Nyquist frequency, fast data transmission is possible without increasing the clock frequency. However, when PAM-4 is applied to a GDDR interface that uses single-ended signaling for high pin efficiency, the maximum eye margin becomes 1/6 VDDQ. Therefore, the maximum eye margin of PAM-4 I/O using 1.2V supply voltage is 200mV. For this reason, in the next-generation GDDR interface, a method of adopting duobinary [5], [61] or 3b/2UI PAM-3 signaling [7] instead of PAM-4 is under consideration.

Quarter-rate clocking scheme and T-coils are conventional circuit elements in recent high speed wirelines, but not in memory interfaces. Fig. 2.3 shows the WCK clocking of GDDR6 [60] . GDDR interface still adopts double data rate, but some products provide optional quad data rate function using PLL. T-coil has not been considered because of the insufficient number of layers of the DRAM process. However, as a method of designing an asymmetric T-coil using an RDL layer has recently been proposed, the possibility of GDDR interface which is faster than 30Gb/s/pin is increasing [9], [48].



Fig. 2.3 WCK clocking of GDDR6 [60]

A training sequence is used to improve the reliability of the GDDR interface. The Fig. 2.4 is a diagram showing the training sequence of GDDR6 [59]. It consists of initialization, WCK2CK alignment, and read & write training, and optional command address (CA) training could be added. This training sequence could be widely applied to transceivers using forwarded clocking as well as GDDR.



Fig. 2.4 Training sequence of GDDR6 [59]

2.3 Wide and Slow Memory Interface

With the advent of silicon interposer technology and TSV, new ways to improve bandwidth have been proposed. Fig. 2.5 is a schematic showing the concept of an embedded multi-die interconnect bridge (EMIB) [62]. By using short and dense interconnects instead of the organic material channel of the PCB, which is commonly used, the data throughput could be dramatically increased. It is suitable for short-reach links such as die-to-die communications and on-chip serial links, and has the advantage of low hardware burden due to low channel loss and short clock distribution network [63] -[84]. Since there are a few companies producing interconnects and the package price is higher than conventional packages, wide and slow interface is applied only to limited applications.



Fig. 2.5 Schematic of the EMIB [62]

Table 2.2 is a summary of prior arts of HBM1, HBM2, HBM2E, and HBM3. Although the pin data rate is less than 10Gb/s, the max bandwidth far exceeds 100GB/s and is expected to exceed 1TB/s in the next generation. Since 3-D stacked memory has high density, HBM is expected to show superior performance than other memories in applications that require a lot of data. Fig. 2.6 shows the structure and cross section of the HBM interface [56]. When the TSV delivers data through the DRAM stack, the 2.5-D silicon interposer deliver data for PHY-to-PHY communication. The length of the silicon interposer may vary depending on the application. For instance, it is known to be around 6mm on average for HBM interface.

Fig. 2.7 shows the configuration of the HBM controller and DRAM transceiver and clocking scheme [85]. HBM also adopts forwarded clocking like other memories, and delay lines are located in the transmitters and receivers to align the clock phase. Although the clock skews between transceivers are not large as the transceivers are very tightly packed, the maximum phase difference of 1000+ I/Os can be large enough to deteriorate signal integrity. Although a training sequence such as GDDR's WCK2CK alignment has not been published as a standard, there is room for its function in HBM's initialization sequence. The Fig. 2.8 is the initialization sequence with lane repairs of the JEDEC standard [85]. If there is a training sequence that can optimize the link for a time less than 1-ms, it is expected that a large number of I/O pins could be foreground or periodically trained.

	HBM1		HBM2		HBN	12E	HBM3
reature	ISSCC'2014 [52]	ISSCC'2016 [53]	ISSCC'2016 [54]	ISSCC'2018 [55]	ISSCC'2020 [56]	ISSCC'2020 [57]	ISSCC'2022 [58]
VDD, VDDQ	1.2V, 1.2V	1.1V, 0.4V					
Data rates	1Gb/s/pin	2.4Gb/s/pin	2.0Gb/s/pin	2.66Gb/s/pin	5.0Gb/s/pin	4.0Gb/s/pin	7.0Gb/s/pin
I/O width	1024	1024	128	1024	1024	1024	1024
Max Bandwidth	128GB/s	307GB/s	256GB/s	341GB/s	640GB/s	512GB/s	896GB/s
Density	8Gb	8Gb+1Gb	8Gb	8Gb+1Gb	16Gb+2Gb+1.5Gb	16Gb	16Gb
Channel count	Ø	16p	8 (16p)	8	8	8	16

Table 2.2
Summary
of HBM1~
HBM3 [52
]-[58]



Fig. 2.6 Illustration of eight-high stacked HBM in a 2.5-D SiP [56]



Fig. 2.7 High Level Block Diagram Example of Clocking Scheme [85]

Fig. 2.8 Initialization Sequence with Lane Repairs [85]



Symbol	Description	Min	Max
t _{INIT0}	Power supply ramp time	0.01	200
tiniti	RESET_n signal LOW time at power-up (after stable power)	200	-
t _{INIT2}	CKE LOW time before RESET_n deassertion	10	-
t _{INIT3}	CKE and WRST_n LOW time after RESET_n deassertion	500	-
t _{INIT4}	Stable clock before CKE HIGH	10	-
t _{INIT5}	Idle time before first MRS command	200	I
tinit6	RDQS_t, RDQS_c driven valid and AERR, DERR driven LOW after RESET_n assertion	I	100
tpw_RESET	RESET_n signal LOW time with stable power	1	1

Chapter 3

Background on Voltage-Mode Drivers for Memory Interfaces

3.1 Overview

Voltage-mode drivers are widely used in memory interfaces because they are more power efficient than current-mode drivers. In DDR, data was driven using the stub series terminated logic (SSTL) method. This method was able to easily sense data by setting the reference voltage to 0.5VDD regardless of conditions. From SSTL25 of DDR1 to SSTL15 of DDR3, there was only a change in operation voltage, but no change in structure [86]. However, SSTL has two critical problems. First, SSTL has a disadvantage in high-speed operation because of its large I/O capacitance. Since both PMOS and NMOS are required for the signal to be transmitted based on 0.5VDD, the parasitic capacitance is greatly increased. Second, power loss due to DC current of on-die termination (ODT). Due to these problems, a pseudo-open drain (POD) structure has been widely adopted since DDR4. POD eliminates unnecessary DC current by removing low termination NMOS from SSTL. Also, DC current can be removed by maintaining data '1' when there is no data transition.



Fig. 3.1 The pull-up and pull-down operation of (a) SSTL and (b) POD

Low Power DDR (LPDDR) used in mobile devices adopts High Speed Un-terminated Logic (HSUL) for lower power consumption. Since HSUL has no termination, signal integrity deterioration occurs due to reflection, but power efficiency is greatly improved because there is no DC current. In LPDDR3, HSUL was adopted for low-speed operation and POD was adopted for high-speed operation. However, in LPDDR4, as the data rate becomes faster, termination logic that compensates for the disadvantages of HSUL or POD is needed. Low Voltage Swing Terminated Logic (LVSTL) replaces pull-up PMOS with NMOS and adopts VSSQ termination. At this time, the pull-up NMOS operates in the saturation region. From LPDDR4X, power efficiency was further improved by lowering VDDQ to 0.6V or less, and in LPDDR5, it is lowered to a minimum of 0.3V.



Fig. 3.2 The pull-up and pull-down operation of (a) HSUL and (b) LVSTL

3.2 Output Impedance

P-over-N drivers have been widely used in high speed wireline transmitters as well as memory interfaces due to their simple structure and low power consumption [87] -[103]. As the gm of the transistor is changed depending on the output voltage, source-series termination (SST), which reduces the variation range of impedance using a series resistor, is mainly used. The sum of the impedance of the driver's transistor and series resistor is equal to the line impedance. If the target impedance changes or the impedance of the transistor changes due to PVT



Fig. 3.3 Two methods for impedance calibration [104]

variations, impedance calibration is required. Fig. 3.3 shows two representative methods to calibrate the output impedance [104]. The first method is to control the number of driver slices. Since the output impedance is inversely proportional to the number of slices, the range is wide and the impedance is predictable, but for fine control, the impedance of each slice is large and the number of slices must be increased, so not only the area increases, but also the parasitic capacitance of the output node increases. Therefore, for finer control, digital impedance controls can be added to the driver's pull-up and pull-down paths. This method has a narrow impedance range comparing to the former method. Actual drivers combine these two methods appropriately to enable both coarse and fine control.

The N-over-N driver, which replaces the pull-up PMOS of the SST driver with an NMOS, has an asymmetrical impedance variation according to the output voltage. Fig. 3.4 shows the change in impedance of pull-up NMOS and pull-down NMOS according to V_{out} under certain conditions. Since the change in resistance of the pull-up NMOS is larger than that of the pull-down NMOS, the difference in



Fig. 3.4 Output impedance of N-over-N driver
impedance occurs asymmetrically depending on the output voltage. This not only impairs signal integrity in NRZ signaling, but also reduces BER when multi-level signaling is applied

A method of improving the RLM of PAM-4 signaling by adding a pull-up NMOS to the N-over-N driver has been proposed [105]. Fig. 3.5 shows the improved PAM-4 driver and its operating principle. The driver uses three pull-up NMOS that operate differently according to the PAM-4 level, enabling multi-level signal-ing efficiently at low VDDQ by adding only one transistor and encoder.



Fig. 3.5 Impedance-matched PAM-4 N-over-N driver and its operation [105]

Chapter 4

HBM Interface with Per-Pin Self-Training sequence

4.1 Overview

With the emergence of through silicon via technology, memory is stacked in the vertical direction. In addition, with the development of silicon interposer technology, a very high-density memory interface is realized. Therefore, interest in area and power consumption of the memory interface is growing. Since thousands of transceivers simultaneously transmit data in HBM, even a small increase in the power consumption of each transceiver greatly increases the power consumption of the entire system. Large power consumption harms the stability of the power distribution network and hinders the operation of the system due to high heat generation. Therefore, the power efficiency of each transceiver greatly affects the performance of the entire system. However, since power consumption is in a trade-off relationship with most parameters, it is important to find an optimum point balance between performance and power consumption.

This paper presents a 68.7-fJ/b/mm 375-GB/s/mm single-ended PAM-4 interface with per-pin training sequence for the next-generation HBM controller [106] . In order to maximize energy efficiency, low 0.6-V VDDQ is utilized, and a charge-recycling sampler is proposed. A self-training sequence incorporating eye optimization, eye centering, and offset calibration is designed to meet BER<10-12, overcoming the disadvantages of the decreased voltage margin and the increased mismatch between the transceiver components due to the low-power design.

4.2 Proposed HBM Interface

4.2.1 Low-Power Single-Ended PAM-4 Transceiver

Fig. 4.1 is a low-power HBM interface that includes a memory interface controller that can apply the proposed training sequence. The HBM interface consists of eight 375-GB/s/mm DQ lanes, a pair of 3-GHz differential clock lanes, and a 375-Mb/s bidirectional sideband. Clock lanes and sideband are shared by 8 DQ lanes. On the controller-side where the transmitter is located, a memory interface controller that controls the training sequence is located, and on the memoryside including the receiver, there is a memory digital block that supports functions for training. At this time, the memory side is composed of only very simple blocks, considering performance of DRAM process. In the DQ lanes of the memory side, only the PAM-4 samplers and PAM-4 decoders are located. In the DQS lanes, a forwarded clock receiver, an IQ divider, and an 8-b VDAC that generates the reference voltages of the samplers are located. Memory digital block supports training sequence by controlling VDAC and sideband transceiver.

On the other hand, the controller-side contains most of the blocks required for the training sequence. In the DQ lane, the pattern generator that creates the PAM-4 DC level, the SBR pattern, and the evaluation pattern is located. The generated patterns are transmitted through the 16:2 serializes, the N-over-N drivers and FFEs. In the DQS lanes, there is an IQ divider that receives external clocks and generates 8-phase clocks, and a phase interpolator that controls the clock phases.



Fig. 4.1 Overall architecture of the proposed HBM interface

The memory interface controller calculates the digital control signal of each block and communicates with the memory-side through the sideband transceiver.

The blocks that play the most key role in the training sequence are the phase interpolator and VDAC. Fig. 4.2 is a schematic of an 8-phase interpolator. The 9b phase interpolator supports 8-UI phase sweep, allowing the overall shape of the SBR pattern to be monitored. After two 4:1 multiplexers in the first stage select the clock phase, the current-mode phase mixer provides 6-b control. In the last stage, the output clock is passed through CML to CMOS logic to convert it to a rail-to-rail swing. Four identical phase interpolators control a total of eight clock phases. Fig. 4.3 is the post-layout simulation results of the phase interpolator. The most important thing in training is to maintain the monotonicity of the phase delay. If the monotonicity is broken when scanning the SBR pattern, the peak point may be detected at the wrong location, which makes the training



Fig. 4.2 Schematic of the 3-b phase selector and 6-b phase interpolator



Fig. 4.3 Post-layout simulation result of the phase interpolator.

sequence inaccurate. Post-layout simulation shows that the monotonicity of the clock phase is well maintained. DNL is kept less than 1 LSB, and the maximum value of INL is 4 LSB. The 16:2 serializer uses the PI's output clock to generate the driver and FFE's inputs. 1-UI pulse is generated using two adjacent phases, and 8 data pulses are serialized with CMOS logic gates, as shown in Fig. 4.4. Since the input of the FFE is 1-UI delayed data of the main driver, the input clock phase is rotated by 1-UI. Unlike the SST driver, the N-over-N driver uses inverted input data together, so the timing is aligned by connecting the inverter and transmission gate to the output of the serializer.



Fig. 4.4 Schematic of the PAM-4 transmitter

4.2.2 Charge-Recycling Latch

The data transmitted through the channel is detected as 0 or 1 by the samplers. For the sampler structure, a combination of a slicing latch that compares the size of data and classifies the output into VDD or VSS and an SR latch that maintains output data has been widely used. The StrongARM latch shown in Fig. 4.5 is the most representative circuit that receives differential input data and creates output data as VDD or VSS. Two cross-coupled pairs detect small differences of the inputs during the decision phase. When the clock enters the HIGH state, it becomes a reset phase, and four reset NMOS discharge each node. As the decision phase and the reset phase are repeated, the input data is determined for each clock edge. The reset phase resets the voltage of each node to a constant value,



Fig. 4.5 Schematic of conventional PMOS StrongARM latch

so the influence of previous data on the decision can be removed. However, since the remaining charge is lost in this process, the decision time becomes longer and the power consumption increases.

Differential split-level (DSL) has been proposed to speed up the operation of high-speed CMOS logic and save power consumption [107] -[116]. The proposed charge-recycling latch replaces the reset transistor of the strongARM latch with a DSL switch, recycling the remaining data, making it energy efficient and reducing the decision time. Fig. 4.6 shows the schematic of the charge-recycling latch. The output pair and nodes X and Y are shorted by NMOS switches when the clock is HIGH and share the remaining charge. Then, this charge is recycled when OUT or OUTB becomes VDD in the next decision phase. Fig. 4.7 shows the voltage of each node according to the operation of the charge-recycling latch. OUT and



Fig. 4.6 Schematic of the proposed charge-recycling latch

OUTB are charged to VDD or discharged to VSS in the decision phase and then pre-charged to the same voltage in the reset phase. Since this pre-charged voltage is lower than the threshold voltage, it does not affect the operation of the SR latch. Nodes X and Y need to be completely discharged and recharged in the reset phase, but they can be reset to a high voltage to save a lot of power. With the help of charge-recycling, the power consumption of the proposed sampler is reduced by 44.5%.



Fig. 4.7 Voltages of output nodes during reset and decision phase

4.3 Proposed Per-Pin Training Sequence

4.3.1 Overall Training Sequence

If all DQ transceivers of the HBM interface are trained sequentially, the total training time will increase by the number of interposers. Therefore, it is more efficient to train a certain number of transceivers in a group in parallel. In the proposed HBM interface, eight transceivers are assumed as one group and sequentially trained, and the total training time is set to less than 1ms. An eye monitor has been used to evaluate link performance during the training process and reflect the results to the training sequence. However, this method has a limitation in that the training time is long because it must perform a 2-D sweep while transmitting a sufficient amount of random data to evaluate the eye. In the proposed training sequence, the training time is greatly reduced by introducing time-efficient DC-based training and SBR-based training.

Fig. 4.8 compares 2-D eye monitoring and SBR-based training. The conventional method of evaluating the eye is to transmit random data and measure the eye opening by sweeping the sampling time and reference voltage of the sampler [117] -[118]. This method can evaluate the eye most accurately, but has the disadvantage that the total number of steps is too large. Therefore, it may be appropriate for eye evaluation after all training is complete, but it is not suitable for use during the training process. 1x2y3x iteration compensates for this shortcoming [119]. This method sweeps the time domain at arbitrary voltage level, then sweeps the voltage level at the center. Then, sweep in the time domain once again at the point expected to be the vertical center. It has the advantage of being able to quickly find the center of the eye and evaluate the eye opening with only 3 sweeps. However, it still requires a sufficient number of random data to generate an eye diagram, and it has limitations in that it is difficult to know whether the current eye opening is the optimal point for training. Also, if the eye is asymmetrical, the difference from the actual eye center may increase.

SBR-based can be an alternative to overcome the above limitations. First, the SBR pattern requires a much shorter training time than random data. Since the insertion loss of the silicon interposer is not large, the long-tail post-cursors are expected to be negligibly small. Therefore, the SBR pattern can be maintained by repeatedly transmitting short UIs. Also, when the sampling point is determined, the size of the precursor and post-cursor can be quickly found by searching at 1-UI intervals. Compared to eye monitoring, it has the advantage of being able to evaluate whether the current





Full-scanning

1x2y3x iteration

SBR-based

	Full-scanning	1x2y3x iteration ¹⁾	SBR-based						
# of time steps		K _t							
# of voltage steps		κ _ν							
# of data samples	К _Р	K _{ul}							
Accuracy	High	Low	Low						
# of test points	K _t K _v	2K _t +K _v	K _t K _v						
# of total steps	K _t K _v K _{PRBS}	(2K _t +K _v)K _{PRBS}	κ _t κ _v κ _{ui}						

Fig. 4.8 Comparison of full-scanning, 1x2y3x iteration, and SBR-based method

transceiver is optimized for the characteristics of the interposer through the amount of remaining cursor.

Fig. 4.9 is a conceptual diagram of the entire training sequence. Training sequence proceeds sequentially for 8 channels. The first step is RLM training that adjusts the DC levels of the PAM-4 signal, and in the case of the first channel, ZQ calibration is performed together. Here, the reference voltage to be used in the receiver can also be determined together. After the DC levels are determined, the sampling point and FFE coefficient are trained while transmitting the SBR pattern. At this time, since the shape of the eye continues to change during the process of optimizing the FFE coefficient, incorrect training is prevented by continuously renewing the reference voltage and sampling point. Finally, the DC levels are transmitted once again and the offsets of the samplers are calibrated. When all processes are finished, repeat this process for the transceiver of the next channel, and when all 8 transceivers are trained, the entire sequence is terminated. All of these processes take less than 1-ms.



Fig. 4.9 Conceptual diagram of the entire training sequence

Except for offset calibration, all training sequences require transmitted PAM-4 DC levels or SBR pattern. When the controller-side transmits the training pattern, the memory-side sweeps the reference voltage to find the signal voltage at the current sampling point. At this time, the training pattern is transmitted in an 8-UI cycle, and the input voltage of the sampler is kept constant because the receiver samples with an 8-phase clock. The memory digital block controls the 8-b VDAC to generate a reference voltage and returns the result to the controller through the sideband. The controller adjusts the PI code, DCDL code, driver strength, FFE coefficient, and pattern generator mode based on the return value. Sideband communication proceeds with half duplex, and communication is performed by adding a header to the current state, requested function or result value.



Fig. 4.10 DQS lanes and the sideband for training

4.3.2 DC-based Training

A training sequence using the DC levels is introduced to calibrate the impedance variation of the N-over-N driver and the offset of the sampler. Transmitting the DC levels not only enables much more accurate sampling than transmitting the data patterns, but also stabilizes the power of the system because the dynamic power of the transmitter is hardly consumed. Therefore, the accuracy of training is improved and the number of iterations can be reduced, so training time is also greatly reduced. Fig. 4.11 is a conceptual diagram of DC-based RLM training. When PAM-4 DC levels are transmitted from the transmitter, the receiver sweeps the reference voltage of the sampler and returns it to the transmitter through the sideband. The controller adjusts the strength of the driver, searches for the optimal point where the value of the PAM-4 DC levels become linear, and determines the output impedance.

Fig. 4.12 shows the N-over-N PAM-4 driver with 2-tap FFE and eye diagram. If the PAM-4 DC levels are V₃, V₂, V₁, and V₀, the size of each eye ΔV_H , ΔV_M , and ΔV_L and reference voltages of samplers can be defined as follow equations.



Fig. 4.11 Conceptual diagram of the DC-based RLM training

$$\Delta V_{\rm H} = V_3 - V_2$$

$$V_{\rm ref, H} = (V_3 + V_2)/2$$
(4.1)

$$\Delta V_{\rm M} = V_2 - V_1$$

$$V_{\rm ref,M} = (V_2 + V_1)/2$$
(4.2)

$$\Delta V_{\rm L} = V_1 - V_0$$

$$V_{\rm ref,L} = (V_1 + V_0)/2$$
(4.3)

To simplify the calculation, let's assume that the driver's supply voltage VDDQ is VDD–Vth. This makes it easy to grasp the tendency of the drain current of NMOS operating in linear mode. Assuming that the ratio of pull-up NMOS and pull-down NMOS is k, and the ratio of main driver and 1-UI delayed FFE driver is α , it can be expressed as follows.



Fig. 4.12 PAM-4 N-over-N driver with 2-tap FFE and PAM-4 DC level

$$V_3 = \left(1 - \sqrt{\frac{k}{k + 1/\alpha}}\right) V_{DDQ} \tag{4.4}$$

$$V_{2} = \left(1 - \sqrt{\frac{k}{k + (\alpha + 2)/(2\alpha + 1)}}\right) V_{DDQ}$$
(4.5)

$$V_{1} = \left(1 - \sqrt{\frac{k}{k + (2\alpha + 1)/(\alpha + 2)}}\right) V_{DDQ}$$
(4.6)

$$V_0 = \left(1 - \sqrt{\frac{k}{k + \alpha}}\right) V_{DDQ} \tag{4.7}$$

Considering the operating range, it can be expected that 0 < k < 1 and $1 > \alpha$. Assume that the supply voltage VDD of the transmitter is 1.0V and the threshold voltage of the transistor is 0.4V. The previous assumption gives VDDQ = 0.6V. If the coefficient of 2-tap FFE is assumed to be about 0.25, $\alpha = 3$. At this time, if the eye height is calculated by changing the pull-up to pull-down ratio k, the result shown in Fig. 4.13 can be obtained. As the k value changes, ΔV_H , ΔV_M , and ΔV_L change with different tendencies, and the RLM ratio also changes. When the pull-down strength is weak, the bottom eye opens wide, but the top eye is small, so the eye does not open efficiently. Conversely, as k increases, the bottom eye gets smaller, but the top eye gets bigger. Since the size of the middle eye changes relatively little, there exists a point where the size of the three eyes becomes similar. The result found through the formula is k = 0.5. The actual simulation result is shown in Fig. 4.14. The size of the three eyes is compared while the pull-down NMOS width is 5-bit digitally controlled (0.2 < k < 0.8). It can be seen that the size of the three eyes becomes the same around k = 0.4. This tendency can also be confirmed through minimum, average, and RLM values of eye openings. Since the minimum opening varies more than the average of the eye openings, the maximum value of the RLM has the same tendency as the minimum opening. Therefore, the optimal strength of the PAM-4 driver can be inferred through the minimum value of eye opening.



Fig. 4.13 Eye height versus pull-up to pull-down ratio



Fig. 4.14 Simulation result of the RLM training

RLM training determines the strength ratio of the pull-up NMOS and pulldown NMOS, but the overall strength is also changed in the process. Perform impedance calibration to match the desired output swing level. Fig. 4.15 is the training blocks used for impedance calibration. The replica pull-up driver is turned on by the controller, and the output level is determined by the ratio of the output impedance and the pull-down resistor's impedance. The sampler senses the DC output level and returns it to the controller. The size of the pull-down resistor can be tuned according to the characteristics of the on-chip channel that mimics the silicon interposer. When the pull-up strength is determined through this process, the controller readjusts the size of the pull-up NMOS and pull-down NMOS of the N-over-N driver according to this value.

The final training using the DC level is the offset calibration of the sampler. Since the low-power sampler has to minimize the amount of current, the size of



Fig. 4.15 Training blocks used for impedance calibration

the transistor is reduced, and the offset increases due to the trade-off. Fig. 4.16 is the sampler's Monte-Carlo simulation results. The standard deviation of the offset is about 17.3 mV, which is quite large. Since the vertical eye opening is less than 100mV because PAM-4 signaling is transmitted with low VDDQ, an offset of 10mV or more may deteriorate BER. Therefore, the offset is calibrated by adjusting the reference voltage of the sampler. In order for training to proceed uniformly, offset calibration is performed at the end after other training is performed with the same sampler. Fig. 4.17 and Fig. 4.18 show training blocks and training methods used for offset calibration. If the transmitter always transmits a constant voltage, the sampler's reference voltage is changed and the result value is output. When the reference voltage is transmitted to the ramp input twice, the output is inverted around the sum of the input voltage and the offset. The memory digital block detects the reference voltage at which the output is inverted and compares it with the sampler used for training. After inversely calculating the offset from the difference between the two values, the offset is added



Fig. 4.16 Monte-Carlo simulation result of the low-power sampler

to the reference voltage determined in the training process and used as a new reference. Through offset calibration, the offset of each sampler is reduced to 2.3mV, the resolution of 8-b VDAC.



Fig. 4.17 Training blocks for offset calibration



Fig. 4.18 Conceptual diagram of offset calibration

4.3.3 SBR-based Training

As mentioned in 4.3.1, an SBR-based training sequence is introduced to significantly reduce the overall training time. The SBR pattern is used for eye centering and FFE coefficient optimization to determine the sampling point of the receiver. Fig. 4.19 shows the training blocks used for eye centering. For the first channel, the phase interpolator searches for the peak point of SBR. For the 2nd to 8th channels, the phase interpolator code is not changed, and the delay difference due to the mismatch between channels is compensated for by adjusting the 3-b DCDL code. This is the point at which the vertical eye opening is expected to be maximum, and is far from the point where $h_{-1} = h_1$ or $h_{-1/2} = h_{1/2}$, which are the conventional CDR lock points. This is due to the characteristics of the on-chip



Fig. 4.19 Training blocks for eye centering

metal channel. Since the on-chip metal channel has RC-dominant characteristics, the eye is asymmetric. Therefore, the point where the height of the eye is maximized is located to the right of the lock point of the conventional CDR. Which of the two points to set as the sampling point is good for BER can be predicted through the structure of the transceiver. Data is transmitted using the N-over-N PAM-4 driver, and the expected vertical eye opening due to the application of FFE is less than 100mV, but since the Nyquist frequency is 3GHz, the horizontal eye opening is relatively roomy. Therefore, vertical opening will have a greater effect on BER than horizontal opening.

After the peak point of the SBR is found, the size of the post or pre cursors are obtained as shown in Fig. 4.20. The data is shifted by 1-UI by changing the selection bit of the MUX of the phase interpolator, and the voltage is measured at the memory-side. The controller sweeps the FFE coefficient and finds the point where the sum of the remaining cursors except for the main cursor is minimized.



Fig. 4.20 Single-bit response before and after FFE coefficient training

4.4 Measurement Result

To verify the proposed training sequence and low-power design, a multi-channel PAM-4 interface is constructed. Fig. 4.22 is a chip photomicrograph fabricated in a 40nm CMOS process. The on-chip metal channel imitating a silicon interposer is located in the center of the chip, the controller-side is located on the left, and the memory-side is located on the right. The phase interpolator, serializer, DCDL, driver, sampler and DEC essential for transceiver operation are compactly laid out to satisfy the narrow channel pitch. The memory controller, memory digital block, RDAC, and pattern generator required to support the measurement and training sequences are laid out relatively large.

The cross section view of the on-chip channel is shown in Fig. 4.22. It is made of top layer metal to minimize resistance, and the width of the channel is 1.5um and the height is 0.85um. A ground shielding layer to prevent crosstalk is inserted between the channels. Since the channel pitch is 4um and the data bandwidth per lane is 12Gb/s/pin, the data throughput is 3Tb/s/mm. A ground plane is also placed under the channel to prevent the effect of noise. The on-chip channel shows RC dominant characteristics and has an insertion loss of -4.53dB at Nyquist frequency 3-GHz.

Fig. 4.23 is a measurement setup to evaluate the performance of the prototype chip. When a 3-GHz differential clock is generated in the Bit Error Ratio Test (BERT, N4903B), the decoded output is output out of the chip. BERT receives this signal as input and measures BER. Power supply is supplied by E361A power



Fig. 4.21 Chip photomicrograph

supply, and 0.6V VDDQ used for N-over-N driver is supplied by B2967A. Since channel is implemented inside the chip, it is not possible to directly see the transmitter's eyes, so analyze the receiver's output with an oscilloscope (MS071604C).



Fig. 4.22 Cross section view and insertion loss of the metal channel



Fig. 4.23 Measurement setup

An eye monitor is used to evaluate the eye diagram before and after the training sequence. Evaluate the eye by performing a 2-D sweep of the phase interpolator of the transmitter and the reference voltage of the receiver using I2C. Fig. 4.24 shows the eye diagram before and after training when VDDQ is 0.6V. The leftmost eye diagram is evaluated using the initial value set based on the simulation results, which is not suitable, and the eye opening is also inefficient. The far right is the PAM-4 eye diagram after training. The sampling point is moved to the point where the vertical eye opening is maximized, and the eye opening is also changed efficiently. The BER bathtub in both cases is shown in Fig. 4.25. For a fair comparison, comparison is made using the sampling point after training as shown in the center eye diagram in Fig. 4.24. As a result, BER < 10^{-12} is not satisfied at the eye center before training, but BER < 10^{-12} is stably achieved at the sampling point after the training sequence.

Fig. 4.26 and Fig. 4.27 show the results of training when VDDQ = 0.55V. The left eye diagram in Fig. 4.26 is the eye diagram before training. Before training, the BER is greater than 10^{-9} at all sampling points due to the insufficient eye margin of the center eye. However, after training, BER < 10^{-12} is achieved at the sampling point as the driver strength and FFE coefficient are properly adjusted.



Fig. 4.24 Eye diagrams before and after training at VDDQ=0.6V



Fig. 4.25 BER bathtub at VDDQ=0.6V



Fig. 4.26 Eye diagrams before and after training at VDDQ=0.55V



Fig. 4.27 BER bathtub at VDDQ=0.55V

Fig. 4.28 is the area and power breakdown. Power consumption is measured when VDDQ=0.6V. For power efficiency, most of the training blocks after the training sequence are turned off, and only essential blocks consume power. Since there is no termination and equalization in the receiver, most of the power is consumed by the transmitter. The N-over-N driver and FFE consume 12.12mW, while the serializer and clock path consume 16.72mW. The receiver consumes only 8.06 mW of power thanks to its simple structure and the help of a charge-recycling sampler. The PRBS for transceiver performance evaluation is generated from the pattern generator and consumes 1.98mW. Finally, the memory digital block remains turned on to detect the training start signal transmitted from the controller. In order to minimize power consumption, only the minimum function is used until the start signal comes in, consuming 0.66mW. Total power consumption is 39.36mW, and energy efficiency is 0.41pJ/b because 8 lanes transmit data at 12Gb/s/pin.

#	Block	Area(umxum)	Pattern Me	mory Digital Block
1	Memory Controller	200x300	Generator (1.98mW)	(0.66mW)
2	Memory Digital Block	150x450		
3	Driver + FFE	45x32		
4	Serializer	95x50	RX	SER+Clock
5	PI	80x80	(8.06mW)	(16.72mW)
6	DCDL	80x225		
7	Pattern Gen.	120x200		
8	Sampler	140x35		
9	Decoder	20x40	(12.12	mvv)
10	RDAC	200x400		

Fig. 4.28 Area and power breakdown

Table 4.1 compares the performance of the prototype chip to state-of-the-art memory interfaces. PAM-4 signaling is adopted and power efficiency is maximized by using low VDDQ. Compared to other transceivers, the lowest energy efficiency is achieved, and BER $< 10^{-12}$ is achieved through an on-chip training sequence. Table 4.2 compares recently announced on-chip signal links. Unlike other links to which NRZ signaling is applied, PAM-4 signaling is applied to achieve high bandwidth per pin of 12Gb/s/pin, and energy efficiency per length with 6mm on-chip metal channel is 68.7fJ/b/mm, which is better performance comparing to other on-chip serial links.

Technology (nm)	Energy Efficiency per length (fJ/b/mm)	Channel Length (mm)	Throughput (Gb/s/um)	Data Rate per pin (Gb/s/pin)	Signaling	
28	91.4	3.5	1.76	18	NRZ	JSSC'16 [74]
65	472	5	5	10	NRZ	JSSC'18 [77]
65	77.2	10	2	10	NRZ	JSSC'18 [78]
65	254	6	8	4	NRZ	ISSCC'20 [10]
28	64.2	6	2	10	Di-code	ISSCC'22 [82]
65	78.8	5.6	11.54	12	NRZ	ISSCC'22 [83]
40	68.7	6	з	12	PAM-4	This work [106]

Table 4.2
Comparison
with
on-chip
serial
links

Monitor / On-chip Refer Calibration Voltage Calibr	Eqaulization 1-Tap Rx D	Technology (nm) 65	Energy 0.56	TRX Area (mm2) 0.033	VDDQ 1.05	Data Rate (Gb/s) 7	Signaling Duobinary	[5]	JSSC'14
ence ation	Ē								
Off-chip Calibration via UART	self-Equalization	28	0.95	0.01	1.2	10	PAM-4 / QAM-16	[6]	ISSCC'16
Internal Eye- Opening Monitor	1-Tap Rx DFE	28	0.95	0.01	0.6	27	PAM-3	[7]	ISSCC'19
In-Situ Channel- Loss Monitor	2-Tap Rx DFE	65	0.97	0.009	1.2	32	PAM-4	[8]	ISSCC'20
On-chip Training Sequence	2-Tap Tx FFE	40	0.41	8x0.018	0.6	12	PAM-4	[106]	This work

Table 4.1 Comparison with memory interfaces

Chapter 5

Low-Power Transmitter with PN-NP driver and T-coil-combined Edge-Boosting Equalizer

5.1 Overview

Although the high-bandwidth memory market is being divided with the advent of HBM and silicon interposer, the demand for fast-and-narrow interfaces is also steadily increasing. In order to increase the bandwidth per pin, a higher frequency clock distribution network, robust equalizers, and a sophisticated driver design are required. This trend causes the ratio of the area and the power consumption occupied by the interface to increase. Therefore, in order to construct a more cost-efficient memory system, low-power and small-area design should be considered as important.

This thesis presents a 32-Gb/s/pin 0.51-pJ/b single-ended resistor-less impedance-matched transmitter with T-coil-combined edge-boosting equalizer in 40nm CMOS technology [120] . This transmitter has two key ideas to overcome aforementioned issues: 1) PN-over-NP driver capable of impedance matching without using a resistor significantly reduces the chip area, and 2) T-coil-based edge-boosting equalizer does not require static current in the non-transition sequence and enables impedance matching at high frequency. In addition, a CMOS clock error corrector is introduced to remove clock phase error and duty-cycle error. As a result, the proposed transmitter achieves the bandwidth of 32 Gb/s maintaining high signal integrity, small area, and low power consumption.

5.2 Proposed PN-NP Driver

As seen in Chapter 3, the output impedance changes depending on the output voltage due to the inherent non-linearity of the voltage-mode driver. Source-se-ries-termination resistor is a passive element in which resistance is kept constant, so the change in total impedance can be minimized, but it has two side effects. First, the area of the source-series resistor greatly increases the area of the entire driver. In particular, when the number of driver slices is increased for impedance calibration and FFE coefficient control, the total area also increases proportionally. The second disadvantage is the increasing transistor size. As the size of the series resistor increases, the linearity is improved, but the size of the transistor must also be increased accordingly. This implies that the size of the pre-driver should be increased as well as making the area of the main driver larger. Accordingly, the area and power consumption of the entire transmitter are increased.

The proposed PN-over-NP driver improved the non-linearity by supplementing the insufficient current of the auxiliary N-over-P pair. Fig. 5.1 shows the basic structure of the PN-over-NP driver. Transistors M1 and M2 receive input data and perform pull-up or pull-down operation. In this case, the transistor operates in the linear region in most situations. Assuming that there are no transistors controlled by Pctrl and Nctrl for fine impedance control, the current flowing through each transistor is as follows.


Fig. 5.1 Proposed PN-over-NP driver operation

$$I_{M1} = k_{M1} \left((VDDQ - V_{th})(VDDQ - V_{out}) - \frac{1}{2} (VDDQ - V_{out})^2 \right)$$
(5.1)

$$I_{M2} = k_{M2} \left((VDD - V_{th}) V_{out} - \frac{1}{2} (V_{out})^2 \right)$$
(5.2)

Here, the quadratic term related to V_{out} causes nonlinearity. Fortunately, these terms are perfectly square, so the insufficient current can be compensated by operating a transistor that receives an appropriate input in the saturation region. Transistors M3 and M4 are N-over-P pairs that receive inverted data as input. If the input data is 'Logic 1', the gate voltage of M4 becomes GND. At this time, if Vout is higher than V_{th}, M4 is turned on in the saturation region and current flows. The M3 works on the same principle. Expressing this as a formula is as follows.

$$I_{M3} = \frac{1}{2} k_{M3} (VDD - V_{out} - V_{th})^2$$
(5.3)

$$I_{M4} = \frac{1}{2} k_{M4} (V_{out} - V_{th})^2$$
(5.4)

If k_{M1} and k_{M3} are equal, the pull-up resistance can be made linear by making the magnitude of VDDQ similar to VDD-V_{th}. Similarly, by introducing a VSSQ greater than 0, the pull-down can also be made linear, but adding a power domain requires a large cost and reduces the signal swing. Therefore, it is not suitable for single-ended interfaces. Instead, a series NMOS that calibrates the impedance performs this role to some extent, and by adjusting the ratio of K_{M2} to K_{M4} , I_{M4} can compensate for I_{M2} . At this time, if V_{out} is close to the ground voltage, the pulldown PMOS (M4) does not turn on, so the output voltage should have an appropriately high value. If VDDQ termination is performed on the receiver, the output voltage does not decrease to ground, so there is a section where M4 is turned on, so this problem does not occur. Fig. 5.2 is the post-layout simulation result showing how the output impedance changes when the output voltage of the PN-over-NP driver changes. The output impedance seems to be well maintained within ± 10 ohm at the target of 50 ohm. Therefore, the PN-over-NP driver enables an area-efficient transmitter because the output impedance is maintained without a source series resistor. Also, unlike the existing SST driver, which requires that the size of the transistor be increased to ensure linearity, this structure maintains 50 ohms with only the transistor, so the size of the transistor can be reduced.



Fig. 5.2 Output impedance of the PN-over-NP driver

Since the size of the driver is small, the size of the previous stages including the pre-driver and serializer is reduced, thereby reducing the size and power consumption of the transmitter.

In addition, the output swing also increases due to the impedance characteristics of this structure. In general SST driver, the difference between DC impedance and AC impedance is not large because the passive resistor occupies a large part of the output impedance. Here, the DC impedance and AC impedance can be defined as follows.

$$Z_{\rm DC} = V_{out} / I_{out} \tag{5.5}$$

$$Z_{AC} = \partial V_{out} / \partial I_{out}$$
(5.6)

In order to easily check the difference between the two impedances, assuming a pull-down driver composed only of NMOS and a series resistor, the DC impedance at this time is as follows when $V_X = V_{out} - I_{out}R_{SST}$.

$$Z_{DC} = R_{SST} + \frac{V_X}{k_{M2} \left((VDD - V_{th}) V_X - \frac{1}{2} V_X^2 \right)}$$

= $R_{SST} + \frac{1}{k_{M2} \left((VDD - V_{th}) - \frac{1}{2} V_X \right)}$ (5.7)

Also, the AC impedance is

$$Z_{AC} = R_{SST} + \frac{1}{k_{M2} ((VDD - V_{th}) - V_X)}$$
(5.8)

When the input data is 'Logic 1', the driver pulls down the current from the output node to drop the output voltage. At this time, the final DC level is determined by the ratio of the DC impedance of the driver to the termination resistor of the receiver. In general, in the case of SST driver, both DC impedance and AC impedance are terminated with values close to 50 ohms, so the DC level is near VDDQ/2. However, the PN-over-NP driver is free from this limitation because it compensates for nonlinearity by using a transistor rather than a passive element. Fig. 5.3 shows the DC impedance and AC impedance of the proposed driver. If the AC impedance is matched to be 50 ohms, the DC impedance is lower than that. Therefore, the voltage level during pull-down is determined to be lower than VDDQ/2, suggesting a larger output swing. Through this, a larger voltage margin can be secured while maintaining impedance matching.



Fig. 5.3 Swing characteristics of the proposed PN-over-NP driver

5.3 Proposed T-coil-combined Edge-Boosting Equalizer

As the data bandwidth per pin required for a memory interface increases, not only receiver but also the transmitter need to compensate for the channel loss. De-emphasis FFE is the most used equalizer structure in transmitter, but it has three major disadvantages when applied to memory interface. First, the larger the ISI removed, the less the signal swing, which is fatal to single-ended signaling transceivers. Second, when implementing FFE in voltage mode driver, static current path is generated and power is wasted when there is no data transition. Finally, the retiming block to generate 1-UI delay consumes a lot of power because it uses a high-speed clock.

In order to overcome these shortcomings, various structures have been proposed. Addition-only FFE (AFFE) AFFE uses an addition-only FFE filter instead of subtraction of conventional FFE to remove unnecessary static current path, resulting in high power efficiency and robustness to coefficient error. [121] . However, since the principle of FFE is the same, the swing is reduced, and the filter that calculates the retiming block and coefficient consumes a lot of power. Edge-boosting equalizer can solve these problems because there is no dc current path and no retiming block [44] . However, since the equalizer is directly connected to the output node through a series capacitor, the output impedance changes depending on the frequency. To solve this problem, the equalizer can be operated only when a transition occurs by introducing an edge-detector [79]. However, the impedance is still reduced at the moment of transition, and additional power consumption is required.

The proposed T-coil-combined edge-boosting equalizer solves this output impedance issue simply by using a T-coil. In a typical ultra-high-speed wireline, a Tcoil or coupled inductors are used to prevent bandwidth reduction due to the ESD protection circuit and output network. However, T-coil had not been applied to the memory interface because there was no low resistance metal suitable for drawing an inductor in the DRAM process and the area occupied by the T-coil was large. As an asymmetric T-coil using only one RDL layer was proposed, it became possible to reduce the effects of parasitic capacitance of both ESD and receiver. Fig. 5.4 shows the T-coil used in the proposed equalizer. Most of the T-coil is designed with thick metal emulating the RDL layer of the DRAM process, and less than 20% is designed with the thinner-metal layer.



Fig. 5.4 Implementation of the T-coil

The proposed equalizer and its RLC equivalent model are depicted in Fig. 5.5 and Fig. 5.6. Node D is the output node of the PN-over-NP driver. The edge-boosting equalizer that receives the same data D_{in} as input is composed of a tristate inverters and a series capacitor, and is connected to the center load tap (C) of the T-coil. This equalizer increases bandwidth by generating short pulses when data transitions occur. The inductance and parasitic capacitance (C_b) of the asymmetric T-coil increase the impedance at high frequencies to maintain the output impedance.



Fig. 5.5 Schematic of symmetric T-coil-based edge-boosting equalizer



Fig. 5.6 RLC equivalent model of the proposed equalizer

Fig. 5.7 is the post-layout simulation result of the output impedance of the transmitter. Compared with the case without T-coil, the impedance variance depending on the frequency is reduced by 47%. This result implies that the output impedance is effectively maintained during transition. Also, it shows that the impedance does not change significantly even when the strength of the equalizer is adjusted.



Fig. 5.7 Output impedance of the transmitter

Fig. 5.8 is an RLC equivalent model expressing three situations for more accurate comparison. (1) assumes that there is only driver and parasitic capacitances of driver, ESD protection circuit, and output pad without equalizer. In this case, this output network could be considered as a 1-pole RC circuit. This structure is typical output network of the transmitters of the memory interfaces. (2) consists of driver, ESD, output pad and T-coil to reduce the effect of ESD capacitance. This is the conventional output network of high-speed wireline transmitters. (3) shows the proposed transmitter including T-coil-combined edge-boosting equalizer. By comparing these three models, the effects of T-coil and equalizer could be compared. A comparison of the output impedances of (1) and (3) is shown in Fig. 5.9. Thanks to the help of the T-coil, it can be seen that the reduction in impedance is small even though the series capacitor of the equalizer is added. At Nyquist frequency, the output impedance is reduced to 40 ohms, which greatly improves channel matching comparing to (1). In addition, by comparing (2) and (3), the effect of equalizer on transmitter bandwidth could be verified. When the edge-boosting equalizer is added, the AC gain at high frequencies is greatly improved. At 48GHz, which is three times the Nyquist frequency, the loss is improved by about 8dB. Therefore, the rise and fall times of output data could be reduced, so that a wider eye margin could be secured.



Fig. 5.8 RLC equivalent model of three situations



Fig. 5.9 Output impedance and gain depending on the frequency

5.4 Circuit Implementation

5.4.1 Overall Architecture

A transmitter with the proposed PN-over-NP driver and T-coil-combined edge-boosting equalizer is depicted in Fig. 5.10. When the 32-b pattern generator receives the divided 1-GHz clock and generates a 32-b parallel data pattern, data goes through a 32:4 serializer and becomes 8-Gb/s quarter-rate data. The proposed transmitter serializes 4-phase 8-Gb/s data into a high-speed 4:1 MUX composed of CMOS logic gates. Finally, the PN-over-NP driver and equalizer transmit data through the channel. The ESD protection circuit is connected to the load tap of the T-coil along with the output node of the equalizer. Input clock buffer receives 16-GHz differential external clock as input. Buffered clock is divided by the IQ divider. The 8-GHz 4-phase clock generated by the IQ divider is phase and duty adjusted by the CMOS-based clock error corrector.





As the data rate rises rapidly, the power consumption of the MUX to generate full-rate data also rises rapidly. Accordingly, 4:1 MUXs that implement simple logic using a transmission gate or NMOS switch have been proposed [122], [123]. However, these structures limit the size of output swing due to IR drop of stacked transistors and switches. Since the PN-over-NP driver requires that the input swing is rail-to-rail like other voltage mode drivers, these structures are not suitable. In the proposed transmitter, 1-UI pulse is generated using AND gate, and then data is serialized using CMOS logic gates as shown in Fig. 5.11. In addition, the output swing is made rail-to-rail through sufficient buffering. This increases power consumption somewhat, but ensures sufficient voltage margin even at low VDDQ.



Fig. 5.11 High-speed 4:1 serializer (MUX)

5.4.2 CMOS-based Clock Error Corrector

The clock distribution network accounts for a significant portion of the transmitter power consumption. As the clock frequency increases, the number of clock buffers increases significantly, reducing power efficiency and making it vulnerable to mismatches and variations. Therefore, a method to increase the power efficiency of the transmitter using a multi-phase clock is being studied. Multi-phase clocking can be effective for clock distribution, but phase error directly exacerbates the eye margin. Therefore, it is important to calibrate the exact positions of the clock edges.

Several methods have been proposed to calibrate the clock duty. The duty can be changed by adjusting the dc level by using a band-pass filter without a loop or by adjusting the edge by using a half-delay lines. Meanwhile, the quadrature phase error can be adjusted using a delay lines, and a phase interpolator is sometimes used when precise adjustment is required. This 2-stage edge correcting is the most intuitive way to adjust both the duty-cycle and the clock phase, but it requires two adaptation logics and occupies a large area.

In the proposed transmitter, an error corrector that directly adjusts the clock edge is proposed. As depicted in Fig. 5.12, the error corrector consists of two coarse stages that adjust the strength of pull-up and pull-down, and a fine stage composed of a current starved inverter. The control code that controls the rise time and fall time of each stage is a 4-bit digital code, and the total control code is 24-bit. The positions of the rising edge and the falling edge are each controlled by 5-bit using the look-up table.



The operating principle of the error corrector is shown in the Table 5.1. The duty and phase change depending on the moving size and direction of the rising edge and the falling edge. For example, if the position of the rising edge is fixed (R=0) and the position of the falling edge is changed, the clock duty changes. If the positions of the rising edge and the falling edge are moved by the same amount, the clock phase changes. Also, the duty and phase could be adjusted at the same time by adjusting the amount at which the two edges are changed.

Fig. 5.13 shows the post-layout simulation result of the clock error corrector. The resolution of each edge control block is 250fs according to the post-layout simulation, and both INL and DNL are less than ± 1 LSB. Fig. 5.14 compares the eye diagram when the clock edge corrector is not operated and when it is activated. The clock edge corrector improves the eye margin by making the eye of the transmitter using a 4-phase clock uniform.

Control Code	Duty	Phase
R > F > 0	Up	Lead
R = 0 > F	Up	-
0 > R > F	Up	Lag
R = F > 0	-	Lead
R = F = 0	-	-
0 > R = F	-	Lag
0 < <mark>R</mark> < F	Down	Lead
R < F = 0	Down	-
0 > R > F	Down	Lag

Table 5.1 Operation principle of the error corrector



Fig. 5.13 Post-layout simulation result of the clock error corrector



Fig. 5.14 Eye diagram (a) with and (b) without clock error corrector

5.5 Measurement Result

The prototype chip is fabricated with a 40nm CMOS process. Fig. 5.15 shows the chip photomicrograph and area breakdown. Since the PN-over-NP driver can perform output impedance matching without a resistor, it occupies only an area of 14um x 38um. The equalizer uses 4 MOM capacitors, but it occupies only 15um x 38um, and is placed right next to the driver. Therefore, it does not significantly affect the overall area of the transmitter. The largest area is the 4:1 serializer which occupies 38um x 16um. Sufficient logic gates and buffers are used to generate driver inputs with rail-to-rail swing. The clock error corrector and buffer used for clock distribution and correction are manufactured using only CMOS gates and occupy only a small area of 18um x 16um. It can be seen that the area of the total transmitter is 44um x 50um, which is very area-efficient. The T-coil is fabricated using the top metal layer which is the thickest metal of the CMOS process and occupied 54um x 50um. In the actual memory interface, since the T-coil must be drawn using only one RDL, there is room for further increase in area. The total area including T-coil is 5008 um².



Fig. 5.15 Chip photomicrograph and area breakdown

The setup for testing the prototype chip is shown in Fig. 5.16. As an external clock, a 16GHz differential clock generated by a signal quality analyzer (MP1800A) is used. Also, a 16GHz synchronous clock or 4GHz auxiliary clock for measurement is also supplied to the oscilloscope (MSO 73304DX) from the signal quality analyzer. In the DUT, the output of the transmitter and the synchronous clock of 8 GHz are generated and input to the oscilloscope. The digital control code that determines the edge phase of the clock error corrector, the output impedance of the driver, the strength of the equalizer, and the data pattern for test communicates with the PC through I2C. Fig. 5.17 shows the characteristics of the supplied external clock. Time interval error (TIE) is 561.48 fs at 16 GHz, and the histogram is distributed up to about 2.5 ps. Fig. 5.18 is the measurement of the s parameter of the 8mm FR-4 trace used in the test with a network analyzer (E7071C). The insertion loss is -5.55dB at the Nyquist frequency, and the return loss is -7.94dB.

Fig. 5.19 shows the eye diagram measured at 12.8Gb/s, 20Gb/s, and 32Gb/s. At 12.8Gb/s and 20Gb/s, a sufficient voltage margin is secured without the help of the equalizer. At 32Gb/s, the vertical eye openings are 87 mV and 114 mV when the equalizer is turned off and on, respectively, and in both cases, the horizontal timing margin is more than 0.5UI.

The power breakdown at 32Gb/s is shown in Fig. 5.20. The total power consumption of the transmitter with equalizer is 16.3mW, and its energy efficiency is 0.51pJ/b. The driver and VDDQ termination consume 1.35mW, the serializer consumes 9.81mW, and the CEC and clock buffer consumes 5.14mW. The details of the power breakdown are calculated from the post-layout simulation results.

Signal Quality Analyzers (Anritsu MP1800A)

B





Fig. 5.16 Measurement setup





Chapter 5. Low-Power Transmitter with PN-NP driver and T-coil-combined Edge-**Boosting Equalizer**



** Measurement result is separated based on post-layout simulation results

Fig. 5.20 Power breakdown

The performance of the proposed transmitter is summarized and compared with other single-ended transmitters for memory interfaces in Table 5.2. The proposed PN-NP driver enables TX impedance matching while reducing the chip area and saving power. The proposed 2-tap edge boosting equalizer offers better power efficiency compared to a conventional FFE, especially in the non-transition sequence. In addition, the output impedance is well-matched during the transition, unlike other types of equalizers.

		10000140	1000000	100000	1000000	10000	000000	100000
		ISSCC'18 [79]	ISSCC'20 [10]	ISSCC'20 [8]	ISSCC'20 [44]	JSSC'21 [105]	ISSCC'22 [121]	JSSC'22 [61]
	Technology	16nm	65nm	65nm	8nm	65nm	28nm	28nm
	lecillology	Finfet	CMOS	CMOS	FinFET	CMOS	LPP	CMOS
D	ata rate [Gb/s]	25	4	32	18	28	20	21
	Signaling	GRS	NRZ	PAM-4	NRZ	PAM-4	NRZ	Duobinary
Subbly	עסס [א]	0 75	2	2	0.85	1.0	<u>,</u>	1.0
voltage	עססמ [א]	0.70	7.1	7.1	1.35	0.6	-	0.8
	Driver type	Charge Pump	Inverter	SST	High-voltage SST	N-over-N	Inverter	SST
Driver	TX equalization	2-tap Edge boosting	2-tap FFE (post + XT)	3-tap FFE (half-rate)	2-tap Edge boosting	2-tap pre-emphasis	4-tap AFFE (pre & post 2)	3-tap FFE (pre & post)
œ Equalizer	No static current during IDLE state	Х	х	х	0	0	0	Х
	Impedance matching during transition	х	Х	0	х	Х	×	0
Energ	y efficiency [pJ/b]	1.17**	0.9	0.97**	N/A	0.58*	1.18	0.67
	Area [mm ²]	0.0102**	0.0027***	**600'0	4.15	0.033	0.00115	0.0072
* Excludes	PRBS generator and 32)-8 corializor	/		-			

**** Area / (# of I/O)

Chapter 6

Conclusion

In this thesis, methods for designing low-power transceivers for memory interfaces are presented. Prior to the design, two trends to improve the bandwidth of the memory interface, a voltage mode driver suitable for low-power design, and termination logics are studied. Based on the results, methods to solve the problems that may occur in the low-power memory interface, especially using low-VDDQ drivers, are presented.

In a wide-and-slow interface using many DQ pins, a training sequence capable of calibrating nonlinearity of N-over-N drivers, DQS skew, FFE coefficient error, and offsets of samplers has been proposed. Training time is greatly reduced by using PAM-4 DC levels and SBR patterns instead of the existing 2-D eye monitor. In addition, the charge-recycling latch improved the power consumption and decision time, so the sampler's power consumption is saved by 44.5% compared to when the strongARM latch is used. With the help of the training sequence and the charge-recycling latch, the proposed HBM interface achieved 68.7-fJ/b/mm, which is the best energy efficiency comparing to that of state-of-the-art memory interfaces and on-chip serial links. This result suggests that low-VDDQ PAM-4 signaling can be applied to next-generation HBM interfaces if there is an appropriate training sequence, unlike HBM so far that has adopted NRZ signaling.

Second, a PN-over-NP driver and a T-coil-combined edge-boosting equalizer are proposed for fast-and-narrow transmitters that require high per-pin bandwidth. Unlike the SST driver, the PN-over-NP driver enables impedance matching without series resistors, so the area and the power consumption of the driver and pre-driver is reduced. In addition, since the edge-boosting equalizer is connected to the center tap of the T-coil, impedance drop does not occur even at high frequencies, so signal integrity is improved. Thanks to the proposed structures of the driver and the equalizer, the proposed transmitter achieves a power efficiency of 0.51pJ/b, which is the best compared to state-of-the-art single-ended transmitters including an equalizer. Area of the transmitter is 5008um2 including T-coil.

The PN-over-NP driver is expected to be applicable not only to single-ended NRZ signaling but also to multi-level signaling. For example, in the case of PAM-4 signaling in which the voltage eye margin is reduced to 1/3, the difference in impedance at each level can deteriorate signal integrity. The PN-over-NP driver is expected to solve these problems because it shows high linearity regardless of voltage level. In addition, since inverted input data is required, power efficiency can be further increased in differential signaling.

Bibliography

- [1] Robert H. Dennard, "Field-effect transistor memory", U.S. Patent No. 3,387,286A, June, 1967.
- [2] Wulf, Wm A., and Sally A. McKee. "Hitting the memory wall: Implications of the obvious," ACM SIGARCH computer architecture news 23.1, pp. 20-24, 1995.
- [3] A. Biswas and A. P. Chandrakasan, "Conv-RAM: An energy-efficient SRAM with embedded convolution computation for low-power CNN-based machine learning applications," 2018 IEEE International Solid - State Circuits Conference - (ISSCC), 2018, pp. 488-490.
- [4] Y. -C. Kwon et al., "25.4 A 20nm 6GB Function-In-Memory DRAM, Based on HBM2 with a 1.2TFLOPS Programmable Computing Unit Using Bank-Level Parallelism, for Machine Learning Applications," 2021 IEEE International Solid- State Circuits Conference (ISSCC), 2021, pp. 350-352.
- [5] S. -M. Lee et al., "An 80 mV-Swing Single-Ended Duobinary Transceiver With a TIA RX Termination for the Point-to-Point DRAM Interface," in IEEE Journal of Solid-State Circuits, vol. 49, no. 11, pp. 2618-2630, Nov. 2014.
- [6] W. -H. Cho et al., "10.2 A 38mW 40Gb/s 4-lane tri-band PAM-4 / 16-QAM

transceiver in 28nm CMOS for high-speed Memory interface," 2016 IEEE International Solid-State Circuits Conference (ISSCC), 2016, pp. 184-185.

- [7] H. Park, J. Song, Y. Lee, J. Sim, J. Choi and C. Kim, "23.3 A 3-bit/2UI 27Gb/s PAM-3 Single-Ended Transceiver Using One-Tap DFE for Next-Generation Memory Interface," 2019 IEEE International Solid- State Circuits Conference - (ISSCC), 2019, pp. 382-384.
- [8] P. -W. Chiu and C. Kim, "22.4 A 32Gb/s Digital-Intensive Single-Ended PAM-4 Transceiver for High-Speed Memory Interfaces Featuring a 2-Tap Time-Based Decision Feedback Equalizer and an In-Situ Channel-Loss Monitor," 2020 IEEE International Solid- State Circuits Conference - (ISSCC), 2020, pp. 336-338.
- [9] D. Lee et al., "A 16Gb 27Gb/s/pin T-coil based GDDR6 DRAM with Merged-MUX TX, Optimized WCK Operation, and Alternative-Data-Bus," 2022 IEEE International Solid- State Circuits Conference (ISSCC), 2022, pp. 446-448.
- [10] H. -G. Ko, S. Shin, J. Oh, K. Park and D. -K. Jeong, "6.7 An 8Gb/s/µm FFE-Combined Crosstalk-Cancellation Scheme for HBM on Silicon Interposer with 3D-Staggered Channels," 2020 IEEE International Solid- State Circuits Conference - (ISSCC), 2020, pp. 128-130.
- [11] S. Mirabbasi, L. C. Fujino and K. C. Smith, "Through the Looking Glass— The 2022 Edition: Trends in solid-state circuits from ISSCC," in IEEE Solid-State Circuits Magazine, vol. 14, no. 1, pp. 54-72, winter 2022.

- [12] T. M. Hollis et al., "Recent Evolution in the DRAM Interface: Mile-Markers Along Memory Lane," in IEEE Solid-State Circuits Magazine, vol. 11, no. 2, pp. 14-30, Spring 2019.
- [13] Ho Young Song et al., "A 1.2 Gb/s/pin double data rate SDRAM with on-dietermination," 2003 IEEE International Solid-State Circuits Conference, 2003.
 Digest of Technical Papers. ISSCC., 2003, pp. 314-496 vol.1.
- [14] Yongsam Moon et al., "1.2V 1.6Gb/s 56nm 6F2 4Gb DDR3 SDRAM with hybrid-I/O sense amplifier and segmented sub-array architecture," 2009 IEEE International Solid-State Circuits Conference - Digest of Technical Papers, 2009, pp. 128-129,129a.
- [15] K. Sohn et al., "A 1.2 V 30 nm 3.2 Gb/s/pin 4 Gb DDR4 SDRAM With Dual-Error Detection and PVT-Tolerant Data-Fetch Scheme," in IEEE Journal of Solid-State Circuits, vol. 48, no. 1, pp. 168-177, Jan. 2013.
- [16] K. Koo et al., "A 1.2V 38nm 2.4Gb/s/pin 2Gb DDR4 SDRAM with bank group and ×4 half-page architecture," 2012 IEEE International Solid-State Circuits Conference, 2012, pp. 40-41.
- [17] K. -N. Lim et al., "A 1.2V 23nm 6F2 4Gb DDR3 SDRAM with local-bitline sense amplifier, hybrid LIO sense amplifier and dummy-less array architecture," 2012 IEEE International Solid-State Circuits Conference, 2012, pp. 42-44.

- [18] S. Shim et al., "A 16Gb 1.2V 3.2Gb/s/pin DDR4 SDRAM with improved power distribution and repair strategy," 2018 IEEE International Solid - State Circuits Conference - (ISSCC), 2018, pp. 212-214.
- [19] D. Kim et al., "23.2 A 1.1V 1ynm 6.4Gb/s/pin 16Gb DDR5 SDRAM with a Phase-Rotator-Based DLL, High-Speed SerDes and RX/TX Equalization Scheme," 2019 IEEE International Solid- State Circuits Conference - (ISSCC), 2019, pp. 380-382.
- [20] Bong Hwa Jeong et al., "A 1.35V 4.3GB/s 1Gb LPDDR2 DRAM with controllable repeater and on-the-fly power-cut scheme for low-power and highspeed mobile application," 2009 IEEE International Solid-State Circuits Conference - Digest of Technical Papers, 2009, pp. 132-133.
- [21] Y. -C. Bae et al., "A 1.2V 30nm 1.6Gb/s/pin 4Gb LPDDR3 SDRAM with input skew calibration and enhanced control scheme," 2012 IEEE International Solid-State Circuits Conference, 2012, pp. 44-46.
- [22] T. -Y. Oh et al., "25.1 A 3.2Gb/s/pin 8Gb 1.0V LPDDR4 SDRAM with integrated ECC engine for sub-1V DRAM core operation," 2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2014, pp. 430-431.
- [23] C. -K. Lee et al., "23.2 A 5Gb/s/pin 8Gb LPDDR4X SDRAM with powerisolated LVSTL and split-die architecture with 2-die ZQ calibration scheme," 2017 IEEE International Solid-State Circuits Conference (ISSCC), 2017, pp.

390-391.

- [24] N. Kwak et al., "23.3 A 4.8Gb/s/pin 2Gb LPDDR4 SDRAM with sub-100μA self-refresh current for IoT applications," 2017 IEEE International Solid-State Circuits Conference (ISSCC), 2017, pp. 392-393.
- [25] H. -J. Kwon et al., "23.4 An extremely low-standby-power 3.733Gb/s/pin 2Gb LPDDR4 SDRAM for wearable devices," 2017 IEEE International Solid-State Circuits Conference (ISSCC), 2017, pp. 394-395.
- [26] S. -M. Lee et al., "23.6 A 0.6V 4.266Gb/s/pin LPDDR4X interface with auto-DQS cleaning and write-VWM training for memory controller," 2017 IEEE International Solid-State Circuits Conference (ISSCC), 2017, pp. 398-399.
- [27] K. C. Chun et al., "A 16Gb LPDDR4X SDRAM with an NBTI-tolerant circuit solution, an SWD PMOS GIDL reduction technique, an adaptive geardown scheme and a metastable-free DQS aligner in a 10nm class DRAM process," 2018 IEEE International Solid - State Circuits Conference - (ISSCC), 2018, pp. 206-208.
- [28] K. -S. Ha et al., "23.1 A 7.5Gb/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," 2019 IEEE International Solid- State Circuits Conference - (ISSCC), 2019, pp. 378-380.

- [29] H. -J. Chi et al., "22.2 An 8.5Gb/s/pin 12Gb-LPDDR5 SDRAM with a Hybrid-Bank Architecture using Skew-Tolerant, Low-Power and Speed-Boosting Techniques in a 2nd generation 10nm DRAM Process," 2020 IEEE International Solid- State Circuits Conference - (ISSCC), 2020, pp. 382-384.
- [30] Y. -H. Kim et al., "25.2 A 16Gb Sub-1V 7.14Gb/s/pin LPDDR5 SDRAM Applying a Mosaic Architecture with a Short-Feedback 1-Tap DFE, an FSS Bus with Low-Level Swing and an Adaptively Controlled Body Biasing in a 3rd-Generation 10nm DRAM," 2021 IEEE International Solid- State Circuits Conference (ISSCC), 2021, pp. 346-348.
- [31] D. -H. Kim et al., "A 16Gb 9.5Gb/S/pin LPDDR5X SDRAM With Low-Power Schemes Exploiting Dynamic Voltage-Frequency Scaling and Offset-Calibrated Readout Sense Amplifiers in a Fourth Generation 10nm DRAM Process," 2022 IEEE International Solid- State Circuits Conference (ISSCC), 2022, pp. 448-450.
- [32] Sang-Bo Lee et al., "A 1.6-Gb/s/pin double data rate SDRAM with wave-pipelined CAS latency control," in IEEE Journal of Solid-State Circuits, vol. 40, no. 1, pp. 223-232, Jan. 2005.
- [33] J. -D. Ihm et al., "An 80nm 4Gb/s/pin 32b 512Mb GDDR4 Graphics DRAM with Low-Power and Low-Noise Data-Bus Inversion," 2007 IEEE International Solid-State Circuits Conference. Digest of Technical Papers, 2007, pp. 492-617.

- [34] S. -J. Bae et al., "A 60nm 6Gb/s/pin GDDR5 Graphics DRAM with Multifaceted Clocking and ISI/SSN-Reduction Techniques," 2008 IEEE International Solid-State Circuits Conference - Digest of Technical Papers, 2008, pp. 278-613.
- [35] R. Kho et al., "75nm 7Gb/s/pin 1Gb GDDR5 graphics memory device with bandwidth-improvement techniques," 2009 IEEE International Solid-State Circuits Conference - Digest of Technical Papers, 2009, pp. 134-135,135a.
- [36] H. -W. Lee et al., "A 1.6V 3.3Gb/s GDDR3 DRAM with dual-mode phaseand delay-locked loop using power-noise management with unregulated power supply in 54nm CMOS," 2009 IEEE International Solid-State Circuits Conference - Digest of Technical Papers, 2009, pp. 140-141,141a.
- [37] T. -Y. Oh et al., "A 7Gb/s/pin GDDR5 SDRAM with 2.5ns bank-to-bank active time and no bank-group restriction," 2010 IEEE International Solid-State Circuits Conference - (ISSCC), 2010, pp. 434-435.
- [38] S. -J. Bae et al., "A 40nm 2Gb 7Gb/s/pin GDDR5 SDRAM with a programmable DQ ordering crosstalk equalizer and adjustable clock-tracking BW," 2011 IEEE International Solid-State Circuits Conference, 2011, pp. 498-500.
- [39] H. -W. Lee et al., "25.3 A 1.35V 5.0Gb/s/pin GDDR5M with 5.4mW standby power and an error-adaptive duty-cycle corrector," 2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2014, pp. 434-435.

- [40] H. -Y. Joo et al., "18.1 A 20nm 9Gb/s/pin 8Gb GDDR5 DRAM with an NBTI monitor, jitter reduction techniques and improved power distribution," 2016 IEEE International Solid-State Circuits Conference (ISSCC), 2016, pp. 314-315.
- [41] M. Brox et al., "23.1 An 8Gb 12Gb/s/pin GDDR5X DRAM for cost-effective high-performance applications," 2017 IEEE International Solid-State Circuits Conference (ISSCC), 2017, pp. 388-389.
- [42] Y. -J. Kim et al., "A 16Gb 18Gb/S/pin GDDR6 DRAM with per-bit trainable single-ended DFE and PLL-less clocking," 2018 IEEE International Solid -State Circuits Conference - (ISSCC), 2018, pp. 204-206.
- [43] K. -D. Hwang et al., "A 16Gb/s/pin 8Gb GDDR6 DRAM with bandwidth extension techniques for high-speed applications," 2018 IEEE International Solid - State Circuits Conference - (ISSCC), 2018, pp. 210-212.
- [44] S. -M. Lee et al., "22.5 An 8nm 18Gb/s/pin GDDR6 PHY with TX Bandwidth Extension and RX Training Technique," 2020 IEEE International Solid- State Circuits Conference - (ISSCC), 2020, pp. 338-340.
- [45] K. Kim et al., "25.1 A 24Gb/s/pin 8Gb GDDR6 with a Half-Rate Daisy-Chain-Based Clocking Architecture and IO Circuitry for Low-Noise Operation," 2021 IEEE International Solid- State Circuits Conference (ISSCC), 2021, pp. 344-346.
- [46] T. M. Hollis et al., "An 8-Gb GDDR6X DRAM Achieving 22 Gb/s/pin With Single-Ended PAM-4 Signaling," in IEEE Journal of Solid-State Circuits, vol. 57, no. 1, pp. 224-235, Jan. 2022.
- [47] J. Kim et al., "A 60-Gb/s/pin single-ended PAM-4 transmitter with timing skew training and low power data encoding in mimicked 10nm class DRAM process," 2022 IEEE Custom Integrated Circuits Conference (CICC), 2022, pp. 1-2.
- [48] H. N. Rie et al., "A 40-Gb/s/pin Low-Voltage POD Single-Ended PAM-4 Transceiver with Timing Calibrated Reset-less Slicer and Bidirectional T-Coil for GDDR7 Application," 2022 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits), 2022, pp. 148-149.
- [49] J. -S. Kim et al., "A 1.2V 12.8GB/s 2Gb mobile Wide-I/O DRAM with 4×128
 I/Os using TSV-based stacking," 2011 IEEE International Solid-State Circuits Conference, 2011, pp. 496-498.
- [50] S. Takaya et al., "A 100GB/s wide I/O with 4096b TSVs through an active silicon interposer with in-place waveform capturing," 2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers, 2013, pp. 434-435.
- [51] Y. J. Yoon et al., "18.4 An 1.1V 68.2GB/s 8Gb Wide-IO2 DRAM with noncontact microbump I/O test scheme," 2016 IEEE International Solid-State Circuits Conference (ISSCC), 2016, pp. 320-322.

- [52] D. U. Lee et al., "25.2 A 1.2V 8Gb 8-channel 128GB/s high-bandwidth memory (HBM) stacked DRAM with effective microbump I/O test methods using 29nm process and TSV," 2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2014, pp. 432-433.
- [53] K. Sohn et al., "18.2 A 1.2V 20nm 307GB/s HBM DRAM with at-speed wafer-level I/O test scheme and adaptive refresh considering temperature distribution," 2016 IEEE International Solid-State Circuits Conference (ISSCC), 2016, pp. 316-317.
- [54] J. C. Lee et al., "18.3 A 1.2V 64Gb 8-channel 256GB/s HBM DRAM with peripheral-base-die architecture and small-swing technique on heavy load interface," 2016 IEEE International Solid-State Circuits Conference (ISSCC), 2016, pp. 318-319.
- [55] J. H. Cho et al., "A 1.2V 64Gb 341GB/S HBM2 stacked DRAM with spiral point-to-point TSV structure and improved bank group data control," 2018 IEEE International Solid - State Circuits Conference - (ISSCC), 2018, pp. 208-210.
- [56] C. -S. Oh et al., "22.1 A 1.1V 16GB 640GB/s HBM2E DRAM with a Data-Bus Window-Extension Technique and a Synergetic On-Die ECC Scheme,"
 2020 IEEE International Solid- State Circuits Conference - (ISSCC), 2020, pp. 330-332.
- [57] D. U. Lee et al., "22.3 A 128Gb 8-High 512GB/s HBM2E DRAM with a

Pseudo Quarter Bank Structure, Power Dispersion and an Instruction-Based At-Speed PMBIST," 2020 IEEE International Solid- State Circuits Conference -(ISSCC), 2020, pp. 334-336.

- [58] M. -J. Park et al., "A 192-Gb 12-High 896-GB/s HBM3 DRAM with a TSV Auto-Calibration Scheme and Machine-Learning-Based Layout Optimization,"
 2022 IEEE International Solid- State Circuits Conference (ISSCC), 2022, pp. 444-446.
- [59] Graphics Double Data Rate 6 (GDDR6) SGRAM Standard, JESD250C, JEDEC, Feb 2021.
- [60] Micron Technology, "TN-ED-03: GDDR6: The Next-Generation Graphics DRAM", Online(Accessed Dec. 10, 2022), Available: https://www.micron.com/-/media/client/global/documents/products/technicalnote/dram/tned03_gddr6.pdf
- [61] D. Kang et al., "A 21-Gb/s Duobinary Transceiver for GDDR Interfaces With an Adaptive Equalizer," in IEEE Journal of Solid-State Circuits, vol. 57, no. 10, pp. 3083-3093, Oct. 2022.
- [62] Beth Keser; Steffen Kroehnert, "Embedded Multi-die Interconnect Bridge (EMIB)," in Advances in Embedded and Fan-Out Wafer Level Packaging Technologies, IEEE, 2019, pp.487-499.
- [63] B. Kim and V. Stojanović, "An Energy-Efficient Equalized Transceiver for

RC-Dominant Channels," in IEEE Journal of Solid-State Circuits, vol. 45, no.6, pp. 1186-1197, June 2010.

- [64] J. -s. Seo, D. Blaauw and D. Sylvester, "Crosstalk-Aware PWM-Based On-Chip Links With Self-Calibration in 65 nm CMOS," in IEEE Journal of Solid-State Circuits, vol. 46, no. 9, pp. 2041-2052, Sept. 2011.
- [65] J. Lee, W. Lee and S. Cho, "A 2.5-Gb/s On-Chip Interconnect Transceiver With Crosstalk and ISI Equalizer in 130 nm CMOS," in IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 59, no. 1, pp. 124-136, Jan. 2012.
- [66] J. Lee, W. Lee and S. Cho, "A 2.5-Gb/s On-Chip Interconnect Transceiver With Crosstalk and ISI Equalizer in 130 nm CMOS," in IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 59, no. 1, pp. 124-136, Jan. 2012.
- [67] M. H. Nazari and A. Emami-Neyestanak, "A 20Gb/s 136fJ/b 12.5Gb/s/μm on-chip link in 28nm CMOS," 2013 IEEE Radio Frequency Integrated Circuits Symposium (RFIC), 2013.
- [68] Y. Liu et al., "A 0.1pJ/b 5-to-10Gb/s charge-recycling stacked low-power I/O for on-chip signaling in 45nm CMOS SOI," 2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers, 2013, pp. 400-401.
- [69] J. W. Poulton et al., "A 0.54pJ/b 20Gb/s ground-referenced single-ended short-haul serial link in 28nm CMOS for advanced packaging applications," 2013 IEEE International Solid-State Circuits Conference Digest of Technical

Papers, 2013, pp. 404-405.

- [70] S. -H. Lee, S. -K. Lee, B. Kim, H. -J. Park and J. -Y. Sim, "Current-Mode Transceiver for Silicon Interposer Channel," in IEEE Journal of Solid-State Circuits, vol. 49, no. 9, pp. 2044-2053, Sept. 2014.
- [71] E. Mensink, D. Schinkel, E. A. M. Klumperink, E. van Tuijl and B. Nauta, "Power Efficient Gigabit Communication Over Capacitively Driven RC-Limited On-Chip Interconnects," in IEEE Journal of Solid-State Circuits, vol. 45, no. 2, pp. 447-457, Feb. 2010.
- [72] S. Höppner et al., "An Energy Efficient Multi-Gbit/s NoC Transceiver Architecture With Combined AC/DC Drivers and Stoppable Clocking in 65 nm and 28 nm CMOS," in IEEE Journal of Solid-State Circuits, vol. 50, no. 3, pp. 749-762, March 2015.
- [73] G. -S. Byun and M. M. Navidi, "A Low-Power 4-PAM Transceiver Using a Dual-Sampling Technique for Heterogeneous Latency-Sensitive Network-on-Chip," in IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 62, no. 6, pp. 613-617, June 2015.
- [74] B. Dehlaghi and A. Chan Carusone, "A 0.3 pJ/bit 20 Gb/s/Wire Parallel Interface for Die-to-Die Communication," in IEEE Journal of Solid-State Circuits, vol. 51, no. 11, pp. 2690-2701, Nov. 2016.
- [75] N. Wary and P. Mandal, "Current-Mode Full-Duplex Transceiver for Lossy

On-Chip Global Interconnects," in IEEE Journal of Solid-State Circuits, vol. 52, no. 8, pp. 2026-2037, Aug. 2017.

- [76] J. You, J. Song and C. Kim, "A 2-Gb/s/ch Data-Dependent Swing-Limited On-Chip Signaling for Single-Ended Global I/O in SDRAM," in IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 64, no. 10, pp. 1207-1211, Oct. 2017.
- [77] D. Wei, T. Anand, G. Shu, J. E. Schutt-Ainé and P. K. Hanumolu, "A 10-Gb/s/ch, 0.6-pJ/bit/mm Power Scalable Rapid-ON/OFF Transceiver for On-Chip Energy Proportional Interconnects," in IEEE Journal of Solid-State Circuits, vol. 53, no. 3, pp. 873-883, March 2018.
- [78] P. -W. Chiu, S. Kundu, Q. Tang and C. H. Kim, "A 65-nm 10-Gb/s 10-mm On-Chip Serial Link Featuring a Digital-Intensive Time-Based Decision Feedback Equalizer," in IEEE Journal of Solid-State Circuits, vol. 53, no. 4, pp. 1203-1213, April 2018.
- [79] J. W. Poulton et al., "A 1.17-pJ/b, 25-Gb/s/pin Ground-Referenced Single-Ended Serial Link for Off- and On-Package Communication Using a Processand Temperature-Adaptive Voltage Regulator," in IEEE Journal of Solid-State Circuits, vol. 54, no. 1, pp. 43-54, Jan. 2019.
- [80] K. McCollough, S. D. Huss, J. Vandersand, R. Smith, C. Moscone and Q. O. Farooq, "11.3 A 480Gb/s/mm 1.7pJ/b Short-Reach Wireline Transceiver Using Single-Ended NRZ for Die-to-Die Applications," 2021 IEEE International

Solid- State Circuits Conference (ISSCC), 2021, pp. 1-3.

- [81] Y. Nishi et al., "A 0.297-pJ/bit 50.4-Gb/s/wire Inverter-Based Short-Reach Simultaneous Bidirectional Transceiver for Die-to-Die Interface in 5nm CMOS," 2022 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits), 2022, pp. 154-155.
- [82] H. Park et al., "A 0.385-pJ/bit 10-Gb/s TIA-Terminated Di-Code Transceiver with Edge-Delayed Equalization, ECC, and Mismatch Calibration for HBM Interfaces," 2022 IEEE International Solid- State Circuits Conference (ISSCC), 2022, pp. 1-3.
- [83] S. Lee, J. Yun and S. Kim, "A 78.8fJ/b/mm 12.0Gb/s/Wire Capacitively Driven On-Chip Link Over 5.6mm with an FFE-Combined Ground-Forcing Biasing Technique for DRAM Global Bus Line in 65nm CMOS," 2022 IEEE International Solid- State Circuits Conference (ISSCC), 2022, pp. 454-456.
- [84] J. Seo, S. Lee, M. Lee, C. Moon and B. Kim, "A 20-Gb/s/pin 0.0024-mm2 Single-Ended DECS TRX with CDR-less Self-Slicing/Auto-Deserialization to Improve Tolerance on Duty Cycle Error and RX Supply Noise for DCC/CDRless Short-Reach Memory Interfaces," 2022 IEEE International Solid- State Circuits Conference (ISSCC), 2022, pp. 1-3.
- [85] High Bandwidth Memory DRAM (HBM1, HBM2), JESD235C, JEDEC, Mar 2021.

- [86] 최정환, "고속 DRAM Interface," 電子工學會誌, vol. 39, no. 7, pp. 20-26, 2012.
- [87] T. O. Dickson, H. A. Ainspan and M. Meghelli, "6.5 A 1.8pJ/b 56Gb/s PAM-4 transmitter with fractionally spaced FFE in 14nm CMOS," 2017 IEEE International Solid-State Circuits Conference (ISSCC), 2017, pp. 118-119.
- [88] C. Menolfi et al., "A 112Gb/S 2.6pJ/b 8-Tap FFE PAM-4 SST TX in 14nm CMOS," 2018 IEEE International Solid - State Circuits Conference - (ISSCC), 2018, pp. 104-106.
- [89] P. Upadhyaya et al., "A fully adaptive 19-to-56Gb/s PAM-4 wireline transceiver with a configurable ADC in 16nm FinFET," 2018 IEEE International Solid - State Circuits Conference - (ISSCC), 2018, pp. 108-110.
- [90] L. Wang, Y. Fu, M. LaCroix, E. Chong and A. C. Carusone, "A 64Gb/s PAM-4 transceiver utilizing an adaptive threshold ADC in 16nm FinFET," 2018 IEEE International Solid - State Circuits Conference - (ISSCC), 2018.
- [91] E. Depaoli et al., "A 4.9pJ/b 16-to-64Gb/s PAM-4 VSR transceiver in 28nm FDSOI CMOS," 2018 IEEE International Solid - State Circuits Conference -(ISSCC), 2018, pp. 112-114.
- [92] M. -A. LaCroix et al., "6.2 A 60Gb/s PAM-4 ADC-DSP Transceiver in 7nm CMOS with SNR-Based Adaptive Power Scaling Achieving 6.9pJ/b at 32dB Loss," 2019 IEEE International Solid- State Circuits Conference - (ISSCC),

2019, pp. 114-116.

- [93] M. Pisati et al., "6.3 A Sub-250mW 1-to-56Gb/s Continuous-Range PAM-4 42.5dB IL ADC/DAC-Based Transceiver in 7nm FinFET," 2019 IEEE International Solid- State Circuits Conference - (ISSCC), 2019, pp. 116-118.
- [94] T. Ali et al., "6.4 A 180mW 56Gb/s DSP-Based Transceiver for High Density IOs in Data Center Switches in 7nm FinFET Technology," 2019 IEEE International Solid- State Circuits Conference - (ISSCC), 2019, pp. 118-120.
- [95] P. -J. Peng, Y. -T. Chen, S. -T. Lai, C. -H. Chen, H. -E. Huang and T. Shih, "6.7 A 112Gb/s PAM-4 Voltage-Mode Transmitter with 4-Tap Two-Step FFE and Automatic Phase Alignment Techniques in 40nm CMOS," 2019 IEEE International Solid- State Circuits Conference - (ISSCC), 2019, pp. 124-126.
- [96] T. Ali et al., "6.2 A 460mW 112Gb/s DSP-Based Transceiver with 38dB Loss Compensation for Next-Generation Data Centers in 7nm FinFET Technology," 2020 IEEE International Solid- State Circuits Conference - (ISSCC), 2020, pp. 118-120.
- [97] B. -J. Yoo et al., "6.4 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," 2020 IEEE International Solid-State Circuits Conference - (ISSCC), 2020, pp. 122-124.

- [98] M. A. Kossel et al., "8.3 An 8b DAC-Based SST TX Using Metal Gate Resistors with 1.4pJ/b Efficiency at 112Gb/s PAM-4 and 8-Tap FFE in 7nm CMOS," 2021 IEEE International Solid- State Circuits Conference (ISSCC), 2021, pp. 130-132.
- [99] M. -A. LaCroix et al., "8.4 A 116Gb/s DSP-Based Wireline Transceiver in 7nm CMOS Achieving 6pJ/b at 45dB Loss in PAM-4/Duo-PAM-4 and 52dB in PAM-2," 2021 IEEE International Solid- State Circuits Conference (ISSCC), 2021, pp. 132-134.
- [100] D. Xu et al., "8.5 A Scalable Adaptive ADC/DSP-Based 1.25-to-56Gbps/112Gbps High-Speed Transceiver Architecture Using Decision-Directed MMSE CDR in 16nm and 7nm," 2021 IEEE International Solid- State Circuits Conference (ISSCC), 2021, pp. 134-136.
- [101] R. Shivnaraine et al., "11.2 A 26.5625-to-106.25Gb/s XSR SerDes with 1.55pJ/b Efficiency in 7nm CMOS," 2021 IEEE International Solid- State Circuits Conference (ISSCC), 2021, pp. 181-183.
- [102] Z. Guo et al., "A 112.5Gb/s ADC-DSP-Based PAM-4 Long-Reach Transceiver with >50dB Channel Loss in 5nm FinFET," 2022 IEEE International Solid- State Circuits Conference (ISSCC), 2022, pp. 116-118.
- [103] N. Kocaman et al., "An 182mW 1-60Gb/s Configurable PAM-4/NRZ Transceiver for Large Scale ASIC Integration in 7nm FinFET Technology," 2022 IEEE International Solid- State Circuits Conference (ISSCC), 2022, pp.

120-122.

- [104] M. Kossel et al., "A T-Coil-Enhanced 8.5 Gb/s High-Swing SST Transmitter in 65 nm Bulk CMOS With \$≪ -\$16 dB Return Loss Over 10 GHz Bandwidth," in IEEE Journal of Solid-State Circuits, vol. 43, no. 12, pp. 2905-2920, Dec. 2008.
- [105] Y. -U. Jeong, H. Park, C. Hyun and S. Kim, "A 28-Gb/s/pin PAM-4 Single-Ended Transmitter with High-Linearity and Impedance-Matched Driver and 3-Point ZQ Calibration for Memory Interfaces," 2020 IEEE Symposium on VLSI Circuits, 2020, pp. 1-2.
- [106] J. -H. Park et al., "A 68.7-fJ/b/mm 375-GB/s/mm Single-Ended PAM-4 Interface with Per-Pin Training Sequence for the Next-Generation HBM Controller," 2022 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits), 2022, pp. 150-151.
- [107] L. C. M. G. Pfennings, W. G. L. Mol, J. J. J. Bastiaens and J. M. F. van Dijk,
 "Differential split-level CMOS logic for subnanosecond speeds," in IEEE Journal of Solid-State Circuits, vol. 20, no. 5, pp. 1050-1055, Oct. 1985.
- [108] Bai-Sun Kong, Joo-Sun Choi, Seog-Jun Lee and Kwyro Lee, "Charge recycling differential logic for low-power application," 1996 IEEE International Solid-State Circuits Conference. Digest of TEchnical Papers, ISSCC, 1996, pp. 302-303.

- [109] Swee Yew Choe, G. A. Rigby and G. R. Hellestrand, "Dynamic half rail differential logic for low power," 1997 IEEE International Symposium on Circuits and Systems (ISCAS), 1997, pp. 1936-1939 vol.3.
- [110] Swee Yew Choe, G. A. Rigby and G. R. Hellestrand, "Half-rail differential logic," 1997 IEEE International Solids-State Circuits Conference. Digest of Technical Papers, 1997, pp. 420-421.
- [111] Seung-Moon Yoo and Sung-Mo Kang, "CMOS Pass-gate No-race Chargerecycling Logic (CPNCL)," 1999 IEEE International Symposium on Circuits and Systems (ISCAS), 1999, pp. 226-229 vol.1.
- [112] Hongchin Lin, Yi-Fan Chen and Hsien-Chih She, "A low-power 3-phase half rail pass-gate differential logic," ISCAS 2001. The 2001 IEEE International Symposium on Circuits and Systems (Cat. No.01CH37196), 2001, pp. 148-151 vol. 4.
- [113] K. Y. Cheung, "CRRDL: a novel charge recovery-recycling differential logic," ISCAS 2001. The 2001 IEEE International Symposium on Circuits and Systems (Cat. No.01CH37196), 2001, pp. 152-153 vol. 4.
- [114] Jungho Lee, Joonbae Park, Byungjoon Song and Wonchan Kim, "Split-level precharge differential logic: a new type of high-speed charge-recycling differential logic," in IEEE Journal of Solid-State Circuits, vol. 36, no. 8, pp. 1276-1280, Aug. 2001.

- [115] A. Abbasian, S. H. Rasouli, J. Derakhshandeh, A. Afzali-Kusha and M. Nourani, "Race-free CMOS pass-gate charge recycling logic (FCPCL) for low power applications," Southwest Symposium on Mixed-Signal Design, 2003., 2003, pp. 87-89.
- [116] K. Limniotis, Y. Tsiatouhas, T. Haniotakis and A. Arapoyanni, "A Design Technique for Energy Reduction in NORA CMOS Logic," in IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 53, no. 12, pp. 2647-2655, Dec. 2006.
- [117] B. Analui, A. Rylyakov, S. Rylov, M. Meghelli and A. Hajimiri, "A 10-Gb/s two-dimensional eye-opening monitor in 0.13-/spl mu/m standard CMOS," in IEEE Journal of Solid-State Circuits, vol. 40, no. 12, pp. 2689-2699, Dec. 2005.
- [118] H. Noguchi, N. Yoshida, H. Uchida, M. Ozaki, S. Kanemitsu and S. Wada, "A 40-Gb/s CDR Circuit With Adaptive Decision-Point Control Based on Eye-Opening Monitor Feedback," in IEEE Journal of Solid-State Circuits, vol. 43, no. 12, pp. 2929-2938, Dec. 2008.
- [119] M. Kim et al., "A 4266 Mb/s/pin LPDDR4 Interface With An Asynchronous Feedback CTLE and An Adaptive 3-Step Eye Detection Algorithm for Memory Controller," in IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 65, no. 12, pp. 1894-1898, Dec. 2018.
- [120] J. -H. Park et al., "A 32-Gb/s/pin 0.51-pJ/b Single-Ended Resistor-less Impedance-Matched Transmitter with a T-coil-Based Edge-Boosting Equalizer in

40nm CMOS," 2023 IEEE International Solid- State Circuits Conference (ISSCC), 2023.

- [121] C. Moon, J. Seo, M. Lee, I. Jang and B. Kim, "A 20 Gb/s/pin 1.18pJ/b 1149µm2Single-Ended Inverter-based 4-tap Addition-Only Feed-Forward Equalization Transmitter with Improved Robustness to Coefficient Errors in 28nm CMOS," 2022 IEEE International Solid- State Circuits Conference (ISSCC), 2022, pp. 450-452.
- [122] J. Kim et al., "A 112Gb/s PAM-4 transmitter with 3-Tap FFE in 10nm CMOS," 2018 IEEE International Solid State Circuits Conference (ISSCC), 2018, pp. 102-104.
- [123] J. Kim et al., "8.1 A 224Gb/s DAC-Based PAM-4 Transmitter with 8-Tap FFE in 10nm CMOS," 2021 IEEE International Solid- State Circuits Conference (ISSCC), 2021, pp. 126-128.

초 록

본 논문은 메모리 인터페이스를 위한 저전력 송수신기를 설계하기 위한 기술들을 제안하였다. 메모리 인터페이스의 대역폭을 향상시키기 위한 두 가지 관점, 즉 채널 간격을 줄여 핀당 대역폭은 느리지만 전체 대역폭을 향상시키는 방향과 고전적인 송수신기와 마찬가지로 핀당 대역폭을 높이 는 방향에서 소비전력을 최소화하는 방안들이 연구되었다.

먼저, 고대역폭 메모리 인터페이스의 전력 소비를 최적화하는 방안이 연구되었다. 많은 수의 송수신기를 최적화하기 위한 훈련과정이 도입되었 다. DC 레벨을 활용한 훈련방법을 통해 드라이버의 출력 강도와 샘플러의 기준 전압 값이 조절된다. 단일 비트 응답(SBR)을 이용한 훈련은 이차원 모니터 보다 훨씬 짧은 시간 동안 클럭 정렬과 보상 계수 최적화가 가능 하게 한다. 제안된 훈련과정을 통해 8 개의 PAM-4 송수신기가 1ms 내에 최적화되어, 낮은 전압에서도 비트 에러율(BER)이 10⁻¹² 이하로 유지된다. 또한 전하-재활용 래치는 샘플러의 전력소비량을 44.5% 절약하고 판단시 간을 줄여 고속 동작이 가능하게 한다. 훈련과정과 전하-재활용 래치의 도움으로 제안된 고대역폭 메모리 인터페이스는 68.7-fJ/b/mm 의 에너지 효율을 달성하였고, 이 결과는 학계 최고 수준의 메모리 인터페이스와 최 근 발표된 칩상 직렬 링크와 비교하여 가장 우수하다.

두번째로, 높은 핀당 대역폭을 전송하는 송신기의 면적과 전력소비량을 최소화하는 방안이 연구되었다. 제안된 PN-over-NP 드라이버는 직렬저항 의 도움 없이 50Ω 정합이 가능하게 하여 드라이버의 면적을 줄이고 드라 이버와 그 이전 단의 소비 전력을 절약한다. 또 T-코일-결합 에지부스팅보 상기는 피드-포워드 보상기의 불필요한 전류낭비를 제거하여 신호 전환이 없을 때 소비 전력을 최소화하면서도 고주파수 대역에서 출력 임피던스를 유지하여 신호 무결성을 향상시킨다. 또한 수동소자를 사용하지 않는 CMOS 기반 클럭 오류 정정기를 사용하여 적은 면적만 사용하여 4-위상클 럭을 효과적으로 보정하였다. 새롭게 제안된 드라이버와 보상기 덕분에 제안된 송신기는 0.51pJ/b 의 전력 효율을 달성하였고, 이는 보상기를 포 함하는 학계 최고수준 단일 종단 송신기들과 비교하여 가장 우수하다. 송 신기의 면적은 T-코일을 포함하여 5008um²이다.

주요어 : 고대역폭 메모리, 그래픽 메모리, 메모리 인터페이스, 칩상 훈 련과정, N-over-N 드라이버, 피드-포워드 보상기, 클럭 정렬, 오차 보정, 단 일 종단, 임피던스 정합, T-코일, 에지부스팅 보상기.

학 번 : 2018-27402