



Master's Thesis

Design of Low power transceiver with post-1 tap cancellation CTLE for DRAM interface

DRAM interface 를 위한 post-1 tap 제거 CTLE 가 포함된 저전력 송수신기의 설계

by

In-Woo Nam

August, 2023

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

Design of Low power transceiver with post-1 tap cancellation CTLE for DRAM interface

지도 교수 최 우 석

이 논문을 공학석사 학위논문으로 제출함 2023 년 8 월

> 서울대학교 대학원 전기·정보공학부 남 인 우

남인우의 석사 학위논문을 인준함 2023 년 8 월

위 원 장 _____ (인) 부위원장 _____ (인)

위 원 (인)

Design of Low power transceiver with post-1 tap cancellation CTLE for DRAM interface

by

In-Woo Nam

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

August, 2023

Committee in Charge:

Professor Su-hwan Kim, Chairman

Professor Woo-Seok Choi, Vice-Chairman

Professor Deog-Kyoon Jeong

Abstract

In the 20 years between the release of the DDR SDRAM standard in 2000 and the release of the DDR5 SDRAM standard in 2020, the bandwidth of DRAM increased by a factor of 40. The rate of bandwidth growth is accelerating, which raises the importance of more space- and power-efficient high-speed interface circuits in memory cells that occupy limited space.

In this paper, a high-speed transmitter and receiver for memory interfaces is proposed. Typically, DFEs used at the receiving end of memory interfaces require additional hardware area and power consumption due to time constraints in the first loop during high-speed operation. In this paper, this problem is solved by proposing a method to remove post-1 cursor through over-equalization of CTLE, and DFE removes only post-2 cursor ISI. In the process, the circuits of the transmitter driver, CTLE, and active inductor were studied.

The transceiver circuit utilizing the CTLE with the proposed active inductor was fabricated in a 28-nm CMOS process and occupied an area of 0.014 mm2. In the operation of transmitting and receiving 10 Gb/s data, the transmitter and receiver combined consumed 25.45 mW of power, and all core circuits maintained stable operation regardless of variations in process, supply voltage, and temperature.

Keywords : memory interface, transceiver, ISI, CTLE, active inductor

Student Number : 2021-21286

Contents

ABS	TRACT		I
CON	TENTS		Π
LIST	r of fig	GURES	IV
LIST	r of tai	BLES	VI
CHA	PTER 1	INTRODUCTION	1
	1.1 Mot	TIVATION	1
	1.2 The	SIS ORGANIZATION	5
СНА	APTER 2	BACKGROUND OF HIGH-SPEED INTERFACE	6
	2.1 Ove	RVIEW	6
	2.2 BAC	KGROUND ON SERIAL LINKS	8
	2.2.1	SERIAL AND PARALLEL COMMUNICATION	8
	2.2.2	CLOCKING SCHEME	10
	2.2.3	OUTPUT IMPEDANCE	10
	2.3 BUII	LDING BLOCKS	13
	2.3.1	Driver	13
	2.3.2	SERIALIZER	15
	2.3.3	DECISION FEEDBACK EQUALIZER (DFE)	17
	2.3.4	CONTINUOUS-TIME LINEAR EQUALIZER (CTLE)	19
	2.3.5	ACTIVE INDUCTOR	22

СНА	PTER 3	DESIGN OF 10 GB/S ELECTRICAL TRANSCEIVER IN	1 28
NM	CMOS		26
	3.1 Des	SIGN CONSIDERATION	26
	3.2 Ove	ERALL ARCHITECTURE	
	3.3 Cire	CUIT IMPLEMENTATION	31
	3.3.1	CLOCK PATH	31
	3.3.2	DRIVER	32
	3.3.3	SERIALIZER	33
	3.3.4	CTLE WITH ACTIVE INDUCTOR	35
	3.3.5	DFE	36
СНА	PTER 4	SIMULATION RESULTS	37
	4.1 Tra	NSIENT ANALYSIS	
	4.2 Pov	VER BREAKDOWN	
	4.3 PER	FORMANCE SUMMARY	40
СНА	PTER 5	CONCLUSIONS	41
BIBI	LIOGRA	РНҮ	42
초 특	<u>=</u> T		50

List of Figures

FIG. 1. 1 STANDARDS OF DDR, LPDDR, GDDR, HBM
FIG. 1. 2 THE DRAM DATA BANDWIDTH GROWTH [2]
FIG. 2. 1 DATA CENTERS GLOBAL IP TRAFFIC GROWTH [6]7
FIG. 2.2 (A) PARALLEL COMMUNICATION AND (B) SERIAL COMMUNICATION9
FIG. 2. 3 ARCHITECTURE OF SERIAL LINKS
FIG. 2. 4 TWO METHODS FOR IMPEDANCE CALIBRATION [27]11
FIG. 2. 5 N-OVER-N DRIVER
FIG. 2. 6 (A) VOLTAGE MODE DRIVER AND (B) CURRENT MODE DRIVER
FIG. 2. 7 SOURCE-SERIES TERMINATED(SST) DRIVER
FIG. 2. 8 CURRENT-MODE LOGIC(CML) DRIVER
Fig. 2. 9 16:1 Serializer architecture
FIG. 2. 10 FREQUENCY DIVIDER
FIG. 2. 11 ARCHITECTURE OF A N-TAP DFE
FIG. 2. 12 CIRCUIT SCHEMATIC OF A CONVENTIONAL SOURCE-DEGENERATED CTLE21
FIG. 2. 13 CONVENTIONAL ACTIVE INDUCTOR: (A) SCHEMATIC AND (B) SMALL-SIGNAL
FIG. 2. 14 2 DIFFERENT PASSIVE SPIRAL INDUCTOR [35]
Fig. 3. 1 Overall block diagram of proposed TRX28
FIG. 3. 2 BLOCK DIAGRAM OF PROPOSED CHIP IN OPERATION
FIG. 3. 3 IMPLEMENTATION OF (A) CML2CMOS AND (B) DCC
FIG. 3.4 (A) DRIVER STRUCTURE AND (B) MAIN SST DRIVER WITH 12 SLICES
FIG. 3. 5 2:1 SERIALIZER ARCHITECTURE
FIG. 3. 6 POST-SIMULATION RESULTS OF 16:1 SER OUTPUT

FIG. 3. 7 SCHEMATIC OF CTLE WITH ACTIVE INDUCTOR	35
FIG. 3. 8 BLOCK DIAGRAM OF 1-TAP DFE	36
FIG. 4. 1 SBR OF CTLE OUTPUT	37
FIG. 4. 2 EYE-DIAGRAM AT (A) CTLE OUTPUT AND (B) TRANSMITTER OUTPUT	38
FIG. 4. 3 POWER BREAKDOWN	39

List of Table

TABLE 4.1	Power breakdown	39
-----------	-----------------	----

Chapter 1

Introduction

1.1 Motivation

DRAM is a type of random-access semiconductor memory that uses a combination of one capacitor and one transistor to store one bit of data. Since IBM's Robert H. Dennard first presented the concept and design of DRAM in 1966 [1], it has undergone many technological advancements over the past 50 years and is now an integral part of computer systems. DRAM has a simple structure compared to SRAM, where six transistors make up a single cell, which allows for high integration, excellent power efficiency, high yield, and low cost. Since the double-data rate synchronous DRAM (DDR SDRAM) standard was released by the Joint Electron Device Engineering Council (JEDEC) in 2000, DRAM has been used in a wide variety of computer system applications. 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. As a result, DRAM is expected to account for more than 50% of the total semiconductor memory market share by 2023.

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 of DDR, LPDDR, GDDR, HBM



Fig. 1. 2 The DRAM data bandwidth growth [2]

Meanwhile, while DRAM bandwidth has been increasing at a rapid pace, the bandwidth required by memory devices is increasing even faster as technologies to handle huge data such as big data, artificial intelligence (AI) technology, and autonomous driving are being developed. Therefore, the DRAM interface is becoming a major bottleneck in computer systems.

In-memory computing (IMC) is a new memory structure in which simple matrix operations are performed inside the memory hardware, and is being studied to alleviate the memory-processor bottleneck of the von Neumann computer architecture. Although it is not yet commercialized for reasons of memory capacity, production cost, and heat generation, a lot of research is being done to implement IMC in various memory form factors such as SRAM [3] and DRAM [4]. One problem with implementing processing-in-memory (PIM) in DRAM is that it is characterized by a DRAM process that is distinct from the general semiconductor logic process. The DRAM process requires the use of materials with high permittivity and large capacitance to improve the refresh cycle, but this acts as a slowdown factor for circuit systems other than memory cells. Therefore, in order to achieve the bandwidth required by the system, it is necessary to design and develop memory I/O circuits that perform high-speed operation under the premise of the memory process.

On the other hand, not only bandwidth but also power efficiency is an important issue in the design of DRAM interface. In general, transceiver bandwidth is in a trade-off relationship with power consumption, so expanding the DRAM interface bandwidth increases the power consumption, heat generation, noise, etc. of the entire system. For example, because the DRAM interface utilizes a large number of DQ pins, high interface power consumption increases simultaneous switching output (SSO) noise, which degrades signal integrity. Furthermore, the heat generated by the interface circuit adversely affects the performance of the DRAM cell and its performance. Therefore, it is also necessary to design the interface to consume low power.

In addition, the area of the memory device is an important factor because it determines the number of chips per wafer, called net die. Therefore, each cell is required to have as little area as possible, and the same is true for the memory interface. On the other hand, the Decision Feedback Equalizer (DFE) used in high-speed receivers suffers from critical path problems when operating at high frequency, so special methods such as loop-unrolling are used to solve this problem in common.

In this thesis, the methodology of designing a high-speed transceiver operating in DRAM will be approached from two perspectives: assuming the design in the DRAM process and how to reduce the area in the receiver.

1.2 Thesis Organization

This thesis is organized as follows. In Chapter 2, the backgrounds of the highspeed interface architectures are presented. The background on serial links including serial/parallel communication, clocking schemes, transmission line, and output impedance is provided. And, the theoretical background about building blocks are presented. The analyses of driver, serializer, continuous-time linear equalizer (CTLE), active inductor, and decision-feedback equalizer (DFE) are discussed.

In Chapter 3, a 10 Gb/s electrical transceiver with post-1 cursor cancellation CTLE is presented. The design consideration on overall transceiver architecture is introduced. Then, the circuit implementation is explained. Specifically, it presents the implementation and simulation results of clock path, driver, serializer, CTLE with active inductor, and DFE.

In Chapter 4, a post-layout simulation results of designed circuit are shown. The data patterns and eye-diagram of DQS and WCK are shown. In addition, it describes the performance of chip like power consumption, area.

Chapter 5 summarizes the proposed works and concludes this thesis.

Chapter 2

Background of High-Speed Interface

2.1 Overview

The data bandwidth required for wireline interfaces has grown exponentially with the rapidly increasing global data traffic as shown in figure 2.1. Correspondingly, DRAM data bandwidth is also increasing rapidly. In 2000, when DDR SDRAM was first commercialized, the per-pin bandwidth of DRAM was only 0.2~0.4 Gb/s/pin, but with the commercialization of DDR5, it has reached 6.4 Gb/s/pin. In particular, GDDR, which is used for graphical applications, requires the highest data bandwidth compared to other DRAM standards, with the GDDR6 interface standard having a bandwidth of 27 Gb/s/pin [5].



Fig. 2. 1 Data centers global IP traffic growth [6]

However, the different characteristics of the DRAM process and logic process are hindering the expansion of DRAM data bandwidth. The fastest announced GDDR6 interface has a bandwidth of sub-30Gb/s/pin, while the fastest recently announced wireline interface has a bandwidth greater than 100Gb/s/pin. In addition, due to the lack of access to actual DRAM processes, academia mainly uses 28-nm CMOS logic processes to mimic the 10-nm DRAM process [7].

In this chapter, we will describe the key characteristics of the 28nm CMOS logic process and the circuits we designed to mimic the DRAM process.

2.2 Background on Serial Links

2.2.1 Serial and Parallel Communication

There are two ways to communicate between chips: parallel communication and serial communication as shown in Fig. 2.2. In the case of parallel communication, multiple data is transmitted through each multiple channels, and it causes clock skews due to transmission to different channels, and it requires many physical channels and causes crosstalk. In addition, since it drives multiple channels, it must consume more power [8]. Therefore, serial communication that serializes and transmits multiple data is generally widely used, and examples include Ethernet, HDMI, PCi Express, and SONET. In this process, a serializer and a deserializer are required. As Fig. 2.3, when parallel data is received from a data source, it is serialized and transmitted. At the receiver, the data is deserialized and passed to the SOC [9]. This process requires a digital block that generates data with a frequency lower than the operating frequency of the interface, where the slow data is serialized and the fast data is transferred to DRAM.



Fig. 2. 2 (a) Parallel communication and (b) serial communication



Fig. 2. 3 Architecture of serial links

2.2.2 Clocking Scheme

There are four main clocking schemes: Synchronous, Mesochronous, Plesiochronous, and Asynchronous. In the case of Synchronous, all chips use the same frequency and phase, and it is mainly used on low-speed buses as the phase difference caused by physical limitations should be negligible. In Mesochronous, the chips share the same frequency but do not share phase information. In this case, phase recovery circuitry such as Clock-Data Recovery (CDR) is required. This is mainly used in fast memory systems, internal system interfaces, MAC/Packet interfaces, etc. In the case of plesiochronous, the clocks share the same frequency, but each clock has a slowly drifting phase. This requires a CDR and is commonly used in highspeed links. Finally, asynchronous clocking is a communication method with no clocks at all, and is mainly used in embedded systems, Unix, Linux, etc. In this thesis, the transceiver was implemented with clock forwarding in a mesochronous clocking manner.

2.2.3 Output Impedance

P-over-N drivers have been widely used in high-speed interfaces and memory interfaces due to their simple structure and low power consumption [10]-[26]. Since the gm of transistors varies with the output voltage level, source-series termination (SST) is widely used to reduce the variation by attaching a serial resistor to the driver output. The sum of the driver transistor and the series resistor equals the output impedance of the driver. If the target impedance varies due to PVT variation, impedance calibration is required. Figure 2.4 shows two typical methods for calibrating the output impedance [27].



Fig. 2. 4 Two methods for impedance calibration [27]

The first method is to divide the transmitter driver into equal small slices to control the number of active drivers. This method has the advantage that the impedance can be predicted well through simulation, but it has the disadvantage that the range of impedance that can be adjusted is small, and the number of slices becomes excessive when the impedance is to be adjusted over a large range. This not only increases the driver area, but also increases the parasitic capacitance of the output node. Therefore, for finer impedance control, digital impedance control must be added to the pull-up and pull-down paths of the driver. This method utilizes transistors for impedance control, which allows for finer control than the first method, but has a narrower impedance control range. Therefore, actual drivers combine these two methods appropriately to implement coarse and fine control.

The N-over-N driver shown in Fig. 2. 5. replaces the pull-up path transistor of the P-over-N driver with an NMOS in PMOS and exhibits asymmetrical impedance variation with output voltage. This is because the resistance change of the pull-up NMOS is larger than that of the pull-down NMOS. This asymmetry not only worsens signal integrity in signaling, but also worsens BER.



Fig. 2. 5 N-over-N driver

2.3 Building Blocks

2.3.1 Driver

When transmitting signals over a transmission line, data must be sent through a driver with an output impedance of 50 Ω to maintain signal integrity with minimal reflections. This driver is the last stage of the transmitter output, and since it is directly connected to the channel, it directly affects the performance (i.e. eye diagram, BER) of the high-speed interface circuit. Broadly speaking, drivers can be categorized into two types: current mode drivers and voltage mode drivers.



Fig. 2. 6 (a) Voltage mode driver and (b) current mode driver

First, the current mode driver transmits data by directly connecting the current source to the transmission line. This driver is relatively less affected by PVT variation because the transistor operates in the saturation region almost all the time and uses a passive resistor as a load device, and it is easy to maintain the output impedance constantly. However, one transistor in a current mode driver's differential pair is always on, so it consumes power continuously. In addition, current mode drivers generally require a larger area because they include a differential pair, which may limit their use in applications where pads are scarce.



Fig. 2. 7 Source-series terminated(SST) driver

Second, voltage mode drivers transmit data by connecting a voltage source to a transmission line. Unlike the current mode driver, which has a constant output impedance due to the load resistor, the transistor of the voltage mode driver is relative-

ly small and has an output impedance that fluctuates depending on the operating situation, and a source-series termination (SST) structure is commonly used to compensate for this [28]. In addition, the termination has the disadvantage of halving the signal swing, which is an advantage in terms of power consumption. This driver has also been adopted as a standard in DRAM memory interfaces constrained by pad number and area because it can implement single-end signaling.



Fig. 2. 8 Current-mode logic(CML) driver

2.3.2 Serializer

A serializer converts low speed multi-bit parallel data into a high speed single data stream. Typical serializers are implemented by stacking 2:1 serializers. For example, a 16:1 serializer consists of four stages of 2:1 serializers [29]. In a half-rate clock system, the input data rate and clock frequency of the 2:1 serializers are the same, and the output data rate is doubled. This means that the number of data streams should halve, and the data rate and clock frequency should double as the serializer progresses through the stages. Therefore, as shown in the figure, the highest frequency clock is input in the last stage, and divided clocks are used in other stages.



Fig. 2. 9 16:1 Serializer architecture



Fig. 2. 10 Frequency divider

2.3.3 Decision Feedback Equalizer (DFE)

The concept of a decision feedback equalizer (DFE) is depicted in Fig. 2.11, in which the postcursor ISI appearing in the uncompensated pulse response can be mitigated by the feedback signal. This compensation is made possible by first making correct decisions on the previously received signals and then correspondingly adjusting the polarity as well as the magnitude of the feedback signal in order to counteract the post-cursor ISI. More specifically, the architecture shown in Fig. 2.11 is known as DFE-FIR in that the feedback path consists of an FIR filter. With an n-tap FIR filter employed in the feedback, an n-tap DFE can be constructed, enabling the compensation for n-tap post-cursor ISI.



Fig. 2. 11 Architecture of a n-tap DFE

Unlike an FFE capable of addressing both pre-cursor and post-cursor ISI, a DFE can only tackle the post-cursor ISI, as a consequence of the need of decoding the previous symbols. Nonetheless, DFEs offer several appealing benefits. For one thing, by mitigating the postcursor ISI, DFEs effectively emphasize the high-frequency signal components, whereas the noise/crosstalk enhancement for FFEs would not be a concern for DFEs, attributed to the digital-level output signals of-fered by well-designed decision circuits. For another thing, DFEs can succeed in compensating for the post-cursor ISI stemming from the reflections, when there are impedance discontinuities existing in the signal path. Especially for the cases where the reflections cause spectral notches, the efficacy of a DFE can be superior to that of an FFE [30].

In light of these advantages, it has been a favorable option to include a DFE in the receiver. Nevertheless, DFE implementations for high-speed operations demand efforts dedicated to meeting the stringent timing constraint. Referring to Fig. 2.11, this DFE architecture falls into the category of direct DFE, where the resolved data signals (i.e., the decisions) are scaled and directly fed back to the summer. The timing delays within this feedback loop lead to the timing constraints for successful post-cursor ISI compensation. For the Nth post-cursor ISI compensation, the timing constraint in a direct DFE design can be expressed as:

$$T_{CKQ} + T_{setup} + T_{mux} < 1 UI$$

where T_{CKQ} is the clock-to-Q delay of the slicer, T_{dhN} is the propagation delay of

the N-th DFE tap, T_{settle} is the settling time of the summer, and T_{setup} is the setup time of the slicer. It can be seen from the equation that the first-tap DFE poses the most stringent timing constraint, which can be the bottleneck in implementing first-tap DFE at high data rates. Typically, we use loop-unrolling, look-ahead, and other methods to circumvent the timing constraint and remove the post 1 cursor. In this thesis, the post-1 cursor will be cancelled by the over-equalized CTLE, and the DFE will remove the post-2 cursor after it is not constrained by the timing constraint.

2.3.4 Continuous-time Linear Equalizer (CTLE)

A continuous time linear equalizer (CTLE) is widely used in high-speed serial links. The concept of a CTLE can be understood as a filter that provides boosted gain at high frequency. This high frequency boost, sometimes referred to as high frequency peaking, can be used to compensate for in band channel loss. Depending on how they are implemented, CTLEs can be categorized into two types: active and passive. A passive CTLE, as the name implies, is a CTLE that is configured to provide peaking at a desired frequency point using only passive elements of resistors, capacitors, and inductors. On the other hand, active devices such as NMOS and PMOS are utilized to form an active CTLE which behaves like an amplifier with its gain peaked at the target frequency. A typical active CTLE structure is shown in Fig. 2.12. This conventional source degenerate CTLE implements peaking by introducing a zero in the transfer function. The location of the zero in the frequency domain, fz, is expressed as follows:

$$f_z = 1/(2\pi R_z C_z)$$

Rz and Cz are the source-degeneration resistance and capacitance. As you can see from the formula, the peaking frequency is determined by the values of Rz and Cz. In addition, the low frequency small signal gain G_{LF} is derived to be:

$$G_{LF} = (g_m R_L)/(1 + \frac{g_m R_z}{2})$$

gm is the transconductance of the transistors, and RL is the resistance of the load resistor. Accordingly, the peaking frequency of this CTLE can be adjusted by changing Rz and Cz, and the change of Rz also changes the gain at low frequency.



Fig. 2. 12 Circuit schematic of a conventional source-degenerated CTLE

There are two design considerations when designing a CTLE. First, the peaking frequency of the implemented CTLE is critical to performance, but it is vulnerable to PVT variation because its value is a term determined by Rz and Cz. Therefore, a tuning mechanism is required after design to compensate for PVT variation. Second, in order to obtain a peaking magnitude as large as desired, low frequency gain must be sacrificed. As its structure suggests, the peaking of a conventional CTLE is

achieved by suppressing the low frequency gain. Therefore, even if the amplifier achieves the largest peaking possible, the low frequency gain may become too small, resulting in a small signal. CTLE can improve the attenuation of the high frequency signal component by effectively improving the overall channel bandwidth. This feature has enabled CTLE stages to be included as receivers in high speed interfaces in the prior art [31][32]. In this thesis, since the driver of the transmitter is driven with a low Vdd for low power consumption, it was necessary for the CTLE to have a low frequency gain with sufficient peaking but not excessively low. Therefore, the use of an active inductor as the load device of the CTLE was considered

2.3.5 Active Inductor

On-chip inductors are typically implemented in passive spiral form as shown in Fig. 2.14. Using a passive spiral form inductor instead of an active inductor has the advantages of lower power consumption, lower voltage headroom requirement, lower noise, and better linearity. However, the implementation of a spiral inductor requires a very large silicon area as it cannot be miniaturized with process scaling. This disadvantage is particularly critical in memory devices where area is critical to the design. Therefore, an active inductor has been proposed as a more area-efficient alternative [33][34]. The most attractive aspect of active inductors is that, unlike passive inductors, they can be implemented using standard logic processes without requiring a large area. Additionally, active inductors have the advantage of being gain tunable, so active inductors can be tuned to compensate for PVT variation.



Fig. 2. 14 2 different passive spiral inductor [35]



Fig. 2. 13 Conventional active inductor: (a) schematic and (b) small-signal

Although there exist different topologies that a shunt-peaking active inductor can assume, their characteristics and principles of operation are very similar. To illustrate how such circuits work, we start by looking at the conventional structure of an active inductor used in the load of a CS amplifier as shown in Fig. 2.13 (a) [36]. The active inductor shown inside the dashed box consists of an NMOS transistor M1 and a resistor RG connected between VDD and the gate of M1. The impedance looking into the active-inductor load is denoted by ZL. Fig. 2.13(b) shows the simplified small-signal model of the active-inductor load. If a test signal Vx is applied to the source of M1, the voltage across Cgs1 is given by

$$V_{gs1} = V_x \frac{\frac{1}{sC_{gs1}}}{R_G + \frac{1}{sC_{gs1}}} = V_x \frac{1}{sR_G C_{gs} + 1}$$

The corresponding current generated that flows into the drain of M1 will be

$$I_{d1} = g_{m1} V_{gs1} = g_{m1} V_x \frac{1}{s R_G C_{gs} + 1}$$

The impedance Z2 looking into the drain, from the source, of M1 can then be obtained by

$$Z_2 = \frac{V_x}{I_{d1}} = \frac{sR_GC_{gs1} + 1}{g_{m1}} = \frac{sR_GC_{gs1}}{g_{m1}} + \frac{1}{g_m}$$

The first term of Z2, i.e., (sRGCgs1)/gm1, is linearly proportional with frequency and is what gives rise to the inductive property of the active-inductor load. Now with a more precise small-signal model of the active-inductor load shown in Fig. 2.5 (a), which includes the drain-to-source conductance gds1 and drain-to-source capacitance Cds1 of M1, the impedance ZL is given by

$$Z_L = \frac{1 + sC_{gs1}R_G}{s^2 R_G C_{ds1} C_{gs1} + s(C_{gs1} + C_{ds1} + R_G C_{gs1} g_{ds1}) + (g_{ds1} + g_{m1})}$$

For frequencies well below the resonance, the impedance looking into the activeinductor load can be approximated by Rs, L, and Cp of an RLC network which has value of:

$$R_{s} = \frac{1}{g_{ds1} + g_{m1}}$$
$$L = \frac{R_{G}C_{gs1}}{g_{ds1} + g_{m1}}$$
$$C_{p} = C_{ds1}$$
$$\omega_{o} = \sqrt{\frac{g_{ds1} + g_{m1}}{R_{G}C_{gs} C_{ds}}}$$

Focus on the value of series resistance Rs of the active inductor depends on both gm1 and gds1 which are related to the drain current and DC bias point of M1 transistor. Additionally, the inductance of the active inductor can be tuned independently by changing the value of RG while keep the others constant.

Chapter 3

Design of 10 Gb/s Electrical Transceiver in 28 nm CMOS

3.1 Design Consideration

This chip was designed with the assumption of driving an off-chip channel, and two transceiver blocks that can function as both transmitter and receiver were implemented to enable data communication using two chips.

Since it is assumed to be implemented on the DRAM process, it was necessary to implement a device environment that is slower than the logic process. Since the details of the DRAM process are confidential to each DRAM manufacturer, we could not get detailed figures, so we proceeded with the core block design using high threshold voltage (hvt) devices, except for the fastest low threshold voltage (lvt) and ultra-low threshold voltage (ulvt) devices in the TSMC 28nm process.

Next, similar to current commercialized DRAMs, the clock is generated internal-

ly and transmitted using differential clock forwarding, and the two datapaths are designed to transmit PRBS data patterns generated from different sources. Clockpath and datapath each have the same design main driver, but the main driver of datapath is designed to function as a termination resistor by controlling the on-off of the SST slice when the chip is operating as a receiver. Therefore, we designed a termination mode using a mux in the stage before the main driver.

The digital blocks included in the chip are I2C [37], PRBS generator [38], and multiplying delay-locked loop (MDLL). Each of these blocks is responsible for the following operations. The I2C block receives control signals from the outside and makes each control signal in the chip enter the appropriate mode for its operation. The PRBS generator is implemented by logic synthesis of the circuit implemented at the register-transfer level (RTL), and there are two blocks for each chip to pass data to the transmitter. The PRBS generator was created so that the chip can be turned off while performing the receiver operation. The MDLL generates the clock signal that is transmitted by the clkpath transmitter. Since this chip is designed assuming a clock forwarding scheme, only the MDLL of the chip operating as a transmitter generates the clock signal, and the chip operating as a receiver distributes the forwarded clock through DCDL, DCC blocks and buffers.

3.2 Overall Architecture

The broadly described entire block diagram of the proposed transceiver is shown in Fig. 3.1. First, the block diagram shown in Fig. 3. 1. (a) is that of a single chip.



Fig. 3. 1 Overall block diagram of proposed TRX

The proposed design is designed so that a single chip can perform either transmitter or receiver functions depending on the mode. Therefore, when performing each operation, unnecessary blocks can be turned off using power gating. However, the main drivers of the transmitters are an exception, as they are made to function as termination resistors for each channel not only in transmitter mode but also in receiver mode, so they are controlled through the mux in the previous stage. The internal structure of the transmitter, receiver, and clockpath receiver roughly depicted in Fig. 3.1 (a) is organized as shown in (b). First, the transmitter uses a low supply voltage of 0.3V from the main driver for low power consumption, so the serializer is not directly connected to the main driver but is designed to be connected through a pre-driver such as a mux. Next, the datapath receiver cancels the post-1 cursor in CTLE and then cancels the post-2 cursor using 1-tap DFE. The CLKPATH receiver consists of a differential amplifier, CML2CMOS, DCDL, and DCC block directly connected to the PAD, and is designed to restore and distribute the forwarded clock signal from the chip that performs the transmitter function. Fig. 3. 2 is a block diagram of the two chips in status of connected and in operation.



Fig. 3. 2 Block diagram of proposed chip in operation

3.3 Circuit Implementation

3.3.1 Clock path

The third diagram in Fig. 3. 2 (b) is a briefly described clockpath. The clock signal passed through the channel is first amplified appropriately by a differential clock amplifier. The clock signal is then amplified to the CMOS level by the CML2CMOS block shown in Fig. 3. 3. (a). The clock signal is then shifted to the appropriate phase by the DCDL and distributed to the chip through the buffer after recovering the duty close to 50% by the Duty Cycle Corrector in Fig. 3. 3. (b). The post-layout



Fig. 3. 3 Implementation of (a) CML2CMOS and (b) DCC

simulation shows that the clock signal that passes through the duty cycle corrector has a duty cycle of 48~52% at all corners considering process, temperature, supply voltage, and RC variation.



3.3.2 Driver

Fig. 3. 4 (a) Driver structure and (b) main SST driver with 12 slices

The transmitter of this chip has a separate datapath and clockpath, but the structure of the main driver is the same. The only difference is that there is a 2:1 mux between the pre-driver and main driver of the clkpath to act as a termination resistor when the chip performs the receiver function. The 16 low-speed bit streams generated by the PRBS generator are converted to a 10Gb/s single data stream through a 16:1 serializer, which is turned into a differential signal to drive the N-over-N driver in the pre-driver and input to the main driver. Another important aspect of the transmitter driving the transmission line is to maintain an output impedance of 50 Ω . For this purpose, a 40 Ω polysilicon resistor, which has a smaller impedance change when operating than a transistor, is used as the SST resistor, and the remaining 10 Ω of impedance is designed to be adjusted by turning the driver slice on and off. Postlayout simulation shows that the output impedance can be maintained near 50 Ω with the DRV_EN signal in 33 pvt corners (process, supply voltage, temperature, RC variation + TT).

3.3.3 Serializer



Fig. 3. 5 2:1 Serializer architecture

The 16-bit digital output signal generated by the PRBS generator is converted to a 1-bit 10 Gb/s signal by a 16:1 serializer. This serializer is designed in the conventional form of a 2:1 serializer stacked as mentioned in Chapter 2.3.2. The sequence of input signals were input to the 16:1 serializer in the order of D[0] to D[15] in order to form the PRBS data pattern. Each 2:1 serializer consisted of latch to align the data to the clock signal and one 2:1 mux, as shown in Fig. 3. 5. The 2:1 mux was simply constructed by connecting two tri-state inverters. The post-layout simulation showed that the eye opening was maintained between 96.1ps and 99.1ps at all corners. (1UI=100ps @ 10Gb/s)



Fig. 3. 6 Post-simulation results of 16:1 SER output

3.3.4 CTLE with Active Inductor



Fig. 3. 7 Schematic of CTLE with active inductor

This transceiver includes CTLE to mitigate both pre-cursor and post-cursor ISI. This is also to prevent the DFE from consuming additional area and power due to critical paths. Therefore, it is very important for this CTLE to eliminate post-1 cursor by over-equalizing. At the same time, it is also important to ensure that post-n cursors (n>2) are not amplified by over-equalizing, since a 1-tap DFE can only remove the 2nd post cursor. Fig. 3. 7 is a CTLE that uses an active inductor as a load device, which was mentioned in Chapters 2.3.4 and 2.3.5. Unlike conventional CTLEs, CTLEs with this structure do not sacrifice low frequency gain in the process of making the peak gain larger, so the reduced receiver input signal can be further

attenuated by using a lower supply voltage for low power consumption. However, since the main driver of the transmitter uses a supply voltage of 0.3V, the common mode voltage is about 75mV, so PMOS was used as the CTLE input device.

CTLE can adjust the resistance of the corresponding mosfet by adjusting the Vresctrl voltage to control the peak gain. However, due to the low input voltage level, a native device was used to secure the operation region of the resistor mosfet.

3.3.5 DFE



Fig. 3. 8 Block diagram of 1-tap DFE

The DFE was implemented as a 1-tap current integrating DFE that removes only the post-2 cursor to avoid additional area and power consumption [39] [40]. Since the DFE processes the signal after the post-1 cursor is removed from the CTLE, a flip-flop is included in the feedback loop for an additional 1 UI delay.

Chapter 4

Simulation Results

4.1 Transient Analysis

Post-layout simulation of the overall transceiver was done by using HSPICE. Single Bit Response (SBR) at CTLE output, Eye-diagram of transmitter output and CTLE output are as follows, respectively.



Fig. 4. 1 SBR of CTLE output



Fig. 4. 2 Eye-diagram at (a) CTLE output and (b) Transmitter output

As shown in Fig. 4.1, the CTLE with active inductor successfully removed the post-1 cursor, and the CTLE output showed an eye opening of more than 65mV at all pvt variation corners, as shown in Fig. 4.2 (a), which crossed the threshold for the current integrating summer of the DFE to determine, and the DFE output was identical to the input PRBS pattern.

4.2 Power Breakdown

Power breakdown was measured through post-layout simulation for 1 cycle of PRBS pattern. The power consumption of each block is as shown in Table 4.1. Figure 4.3 shows the portion of each block occupies in total power consumption.

Subblock	Mode	Power	Ratio
CLKPATH	ТХ	9.67mW	49%
	RX	4.43mW	18%
DATAPATH	ТХ	2.68mW	11%
	RX	8.36mW	33%

Table 4. 1 Power breakdown



Fig. 4. 3 Power breakdown

4.3 Performance Summary

The proposed 10Gbps transceiver with post-1 tap cancellation CTLE for DRAM interface occupied area of 0.014mm². The total power consumption was 25.45mW.

Chapter 5

Conclusions

In this paper, we propose a transceiver system for DRAM interfaces. As the demand for higher DRAM bandwidth grows, the need for power and area efficient DRAM interface circuits will increase. The proposed transceiver eliminates the post-1 cursor ISI by over-equalization using active inductors to avoid the increase in area and power consumption due to the critical path problem of DFE at high frequencies. After layout, simulation results showed that the transceiver had a power consumption of 25.45 mW. The proposed transceiver is fabricated in a 28 nm CMOS process and occupies an area of 0.014 mm².

Bibliography

- Robert H. Dennard, "Field-effect transistor memory", U.S. Patent No. 3,387,286A, June, 1967.
- [2] 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.
- [3] A. Biswas and A. P. Chandrakasan, "Conv-RAM: An energy-efficient SRAM with embedded convolution computation for low-power CNNbased 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] JEDEC. (2021). GRAPHICS DOUBLE DATA RATE 6 (GDDR6) SGRAM STANDARD [Online] [Accessed on 24th May, 2023] https://www.jedec.org/standards-documents/docs/jesd250c
- [6] Cisco, Cisco Visual Networking Index: Forecast and Methodology,

2016-2021 [Online], [Accessed on 24th May] <u>http://www.invest-</u> data.com/eWebEditor/uploadfile/2018052210064650683279.pdf

- 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), Newport Beach, CA, USA, 2022, pp. 1-2, doi: 10.1109/CICC53496.2022.9772814.
- [8] Wikipedia, serial communication. [Online] [Accessed on 24th May, 2023] <u>https://en.wikipedia.org/wiki/Serial_communication</u>
- [9] Texas Instruments. (2016). Keystone II Architecture Serializer/Deserializer (SerDes): User Guide. Texas, MA: Author
- [10] 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.
- [11] 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.
- [12] 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.

- [13] 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.
- [14] 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.
- [15] 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.
- [16] 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.
- [17] 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.

- [18] 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.
- [19] 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.
- [20] 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.
- [21] 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.
- [22] 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.

- [23] 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.
- [24] 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.
- [25] 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.
- [26] 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.
- [27] 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.
- [28] C. Menolfi et al., "A 16Gb/s Source-Series Terminated Transmitter in 65nm CMOS SOI," 2007 IEEE International Solid-State Circuits Con-

ference. Digest of Technical Papers, San Francisco, CA, USA, 2007, pp. 446-614, doi: 10.1109/ISSCC.2007.373486.

- [29] D. Gong, et al. "Development of A 16: 1 serializer for data transmission at 5 Gbps." In Topical workshop on electronics in particle physics, pp. 21-25, Sep. 2009.
- [30] S. Kiran, S. Cai, Y. Zhu, S. Hoyos and S. Palermo, "Digital Equalization With ADC Based Receivers: Two Important Roles Played by Digital Signal Processing in Designing Analog-to-Digital-Converter-Based Wireline Communication Receivers," in IEEE Microwave Magazine, vol. 20, no. 5, pp. 62-79, May 2019, doi: 10.1109/MMM.2019.2898025
- [31] J. Im et al., "6.1 A 112Gb/s PAM-4 Long-Reach Wireline Transceiver Using a 36-Way Time-Interleaved SAR-ADC and Inverter-Based RX Analog Front-End in 7nm FinFET," 2020 IEEE International Solid-State Circuits Conference - (ISSCC), 2020, pp. 116-118, doi: 10.1109/ISSCC19947.2020.9063081.
- [32] M. Erett et al., "A 2.25pJ/bit Multi-lane Transceiver for Short Reach Intra-package and Inter-package Communication in 16nm FinFET," 2019 IEEE Custom Integrated Circuits Conference (CICC), 2019, pp. 1-8, doi: 10.1109/CICC.2019.8780221.
- [33] E. Sackinger and W. Fischer, "A 3 GHz, 32 dB CMOS Limiting Amplifier for SONET OC-48 Receivers," IEEE International Solid-State Circuits Conference (ISSCC), Digest of Technical Papers, pp. 158-159, 2000.
- [34] Chia-Hsin Wu, Jieh-Wei Liao, and Shen-Iuan Liu, "A 1V 4.2mW Ful-

ly Integrated 2.5Gb/s CMOS Limiting Amplifier using Folded Active Inductors," Proceedings of the International Symposium on Circuits and Systems (ISCAS), pp. I-1044-7, 2004.

- [35] R. Murakami, et al. "An ultra-compact LC-VCO using a stacked-spiral inductor." IEICE Electronics Express IEEE Symp. VLSI Circuits Digest of Technical Papers. 2009.
- [36] T.H. Lee, The Design of CMOS Radio-Frequency Integrated Circuits, Cambridge University Press, 2003.
- [37] Wikipedia, I2C. [Online] [Accessed on 24th May, 2023] https://en.wikipedia.org/wiki/I%C2%B2C
- [38] Wikipedia, Pseudorandom binary sequence. [Online] [Accessed on 24th May, 2023] https://en.wikipedia.org/wiki/Pseudorandom binary sequence
- [39] M. Park, J. Bulzacchelli, M. Beakes and D. Friedman, "A 7Gb/s 9.3mW 2-Tap Current-Integrating DFE Receiver," 2007 IEEE International Solid-State Circuits Conference. Digest of Technical Papers, San Francisco, CA, USA, 2007, pp. 230-599, doi: 10.1109/ISSCC.2007.373378.
- [40] G. R. Gangasani et al., "A 16-Gb/s Backplane Transceiver With 12-Tap Current Integrating DFE and Dynamic Adaptation of Voltage Offset and Timing Drifts in 45-nm SOI CMOS Technology," in IEEE Jour-

nal of Solid-State Circuits, vol. 47, no. 8, pp. 1828-1841, Aug. 2012, doi: 10.1109/JSSC.2012.2196313.

초 록

2000 년 발표된 DDR SDRAM standard 로부터 2020 년 발표된 DDR5 SDRAM standard 까지 20 년 간 DRAM 의 대역폭은 40 배 증가하였다. 대역 폭 증가 속도는 점점 빨라지고 있으며, 이에 따라 한정된 공간을 점유하 는 메모리 셀에서 보다 공간, 전력 효율적인 고속 인터페이스 회로의 중 요성이 대두되고 있다.

이 논문에서는 메모리 인터페이스를 위한 고속 송수신기를 제안하였다. 일반적으로 메모리 인터페이스 수신단에서 사용되는 DFE 는 고속 동작 시 첫 번째 루프의 시간 제약 문제로 추가적인 하드웨어 면적 및 전력 소 모를 요구한다. 논문에서는 이 문제점을 CTLE 의 over-equalization 을 통해 post-1 cursor 를 제거하고, DFE 는 post-2 cursor ISI 만을 제거하는 방식을 제 안함으로써 해결하고자 하였다. 그 과정에서, 송신기 driver, CTLE, active inductor 등의 회로에 대한 연구가 진행되었다.

제안된 active inductor 를 포함한 CTLE 를 이용하는 송수신기 회로는 28-nm CMOS 공정으로 제작되었으며 0.014mm² 의 면적을 차지하였다. 10Gb/s 데이터를 송수신하는 동작에서 송신기, 수신기를 합해 25.45mW 의 전력을 소모하였으며, 모든 핵심 회로들이 공정, 공급전압, 온도 등의 변 동에도 관계 없이 안정적인 동작을 유지하였다.

주요어 : 메모리 인터페이스, 송수신기, ISI, CTLE, active inductor 학 번 : 2021-21286