Next Generation Memory Interfaces

Kalika Saxena

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2015-115
http://www.eecs.berkeley.edu/Pubs/TechRpts/2015/EECS-2015-115.html

May 14, 2015

Acknowledgement

I take this opportunity to express my gratitude to my advisors, professors Elad Alon and Vladimir Stojanovic for their able guidance and supervision. I am highly indebted to them for making this project an immense learning experience for me.

I am grateful to my teammates - Kyle James Dillon, Sinan Liu, Miron Veryanskiy and Chenyang Xu for their collaboration and contributions to the project.

University of California, Berkeley College of Engineering

MASTER OF ENGINEERING - SPRING 2015

Electrical Engineering and Computer Science

Physical Electronics and Integrated Circuits

NEXT GENERATION MEMORY INTERFACES

KALIKA SAXENA

This **Masters Project Paper** fulfills the Master of Engineering degree requirement.

Approved by:

1. Capstone Project Advisor:

Signature: _____ Date _____

Print Name/Department: Elad Alon/ EECS

2. Faculty Committee Member #2:

Signature: _____ Date _____

Print Name/Department: Vladimir Stojanovic/ EECS

ABSTRACT

The DDR4 standard allows for memory to be accessed at twice the data rate of its preceding standard, DDR3 while simultaneously reducing the total power consumption and increasing the memory density. In this project a physical layer (PHY) interface for the DDR4 subsystem has been conceptualized, and designed. As memory devices eternally seek to be faster, denser and extremely low power consuming systems, the PHY interface will get us another step further in this quest. In this report, the challenges and circuit techniques associated with the design of this high data-rate interface have been discussed along with the main differentiating factors between DDR4 and its predecessor, DDR3. The design process and results for the receiver front-end of the physical layer are among the major focus points of this report.

Table of Contents

List of Figures and Tables

TEAM PAPERS

1. Project Context and Introduction

Our project, 'Next Generation Memory Controllers' aims to develop a novel memory controller to deliver a high bandwidth interface with the latest memory standard i.e. DDR4 SDRAM (Double Data Rate 4th generation, Synchronous Dynamic Random Access Memory). The DDR4 standard allows for memory to be accessed at twice the data rate of its preceding standard, DDR3 while simultaneously reducing the total power consumption and increasing the memory density. As memory devices eternally seek to be faster, denser and extremely low power consuming systems, our interface will get us another step further in this quest.

In September 2012, JEDEC Solid State Technology Association, the organization that defines standards for the semiconductor industry, released the JESD79-4 for DDR4 SDRAM. With the release of this standard, companies such as Samsung, Micron Technology, Agilent Technologies and several others are developing SDRAMs compliant with this standard. The intent of our project is not to compete with the industry, but to explore the research and development opportunity presented by the new standard. To achieve the higher speed and low power requirements of DDR4, several changes are required at the architecture and circuit level. We aim to implement the strategies required to design a functional interface that meets the specifications of the DDR4 standard.

SDRAM is one of the most commonly used types of memory in computing systems. It is a volatile memory that requires periodic refreshing to store the data. Owing to the speed and structural simplicity, DRAMs are often used as the main memory in personal computers and workstations. DDR is a class of memory that transfers data twice as fast

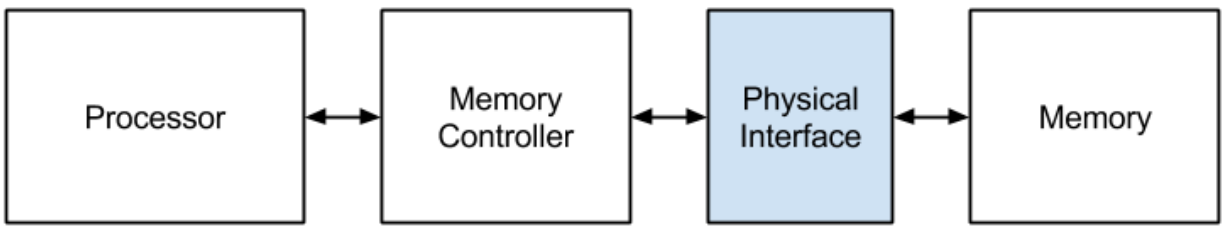

Figure 1 Simplified flow diagram of data interfaces between the processor and memory in a typical computer system

as a SDR (Single Data Rate) memory since the transmission occurs on both positive and negative edge of the clock. At high data rates, this requires careful control of timing margins for reliable operation. DDR4 is the fourth generation of DDR SDRAM which operates at a supply voltage of 1.2 V with data rates up to 3.2 Mbps – twice that of the preceding DDR3 standard. Our project focuses on the design of the physical layer interface that facilitates communication between the DDR4 module and its controller.

2. Industry Analysis and Market Trends

As discussed earlier, we are designing the physical layer interface that allows the data transmission between the memory and the controller. This involves both system and circuit level design of building blocks of the physical layer. This section discusses the current industry and competitive landscape for our project, and enumerates the potential customers, suppliers, as well as new and existing players in this space. A proposal for a potential strategy for bringing it to market is also made keeping these different forces in mind.

Integrated circuit design is an important sector in the semiconductor industry. The industry is known to be highly competitive in nature, and the trend has been increasing over the years (Ulama 2014:19). Product life cycles are short, as more technologically advanced products replace older ones. Adoption of products is majorly affected by performance and reliability. Among the major companies in the memory technology industry is Samsung Electronics (IBISWorld 2015: 27). Broadcom Corporation, Texas Instruments Inc., Advanced Micro Devices (AMD) Inc., Micron Technology Inc. are few of the other major companies that are part of the semiconductor industry (Ulama 2014:19). Noticeably, the major companies in this industry are all fairly large and well established, and compete over products and technologies. High demand for products and extremely low pricing intensify the competition in the industry. This poses a significant barrier to entry for new and smaller companies leading to only several companies currently building DDR4 memory chips and controllers.

As we strive to make a smaller, faster, and more efficient memory interface, we have to compare with the products and research and development efforts of competing companies. Our competitive landscape does not merely include semiconductor companies, but also technologies that have similar features and functions as our project. Existing memory technologies, such as DDR3, 3D stacked (3DS) - DDR3, and GDDR4, compete with DDR4 on various parameters such as cost, speed, and use-cases. While DDR4 is faster than previous memory generations, the higher cost of the new chip technology would make the cheaper DDR3 technology a strong competitor. However, we emphasize the significant performance improvements that DDR4 presents over DDR3 technology which weaken the competition. The following table shows a brief comparison of the key features between two technologies.

Table 1 Comparison between DDR3 and DDR4

Feature	DDR3	DDR4
Power supply voltage	1.5V	1.2 V
Speed	1.6~2.1 Gbps	1.6~3.2 Gbps
Density	8GB(max)	16GB(max)
Price	~ $100	~ $200

The first comparison between DDR3 and DDR4 is from the perspective of power efficiency. Not only does DDR4 have a lower supply voltage, but it also implements a new algorithm to control its energy consumption by entering a “standby” mode more frequently and precisely than DDR3. These improvements lead to better performance in terms of the power consumption, and operating temperature.

One of the most essential features, memory speed, has been improved significantly in DDR4. The analogy between the memory speed and highway traffic speed is very descriptive. The speed of the memory is the amount of data can be transferred in a certain period of time. There are two factors determining the speed, which are, interface width

and frequency of the memory's operation. Considering the analogy, the bandwidth is the quantity of lanes on a highway, and frequency is the travel speed of its vehicles. Within a fixed time period, having more lines and a faster speeds will allow for more vehicles to travel. Similarly, having an improved working frequency, along with an enlarged bandwidth, DDR4 achieves a data transmission speed which is approximately 1.5 times faster than DDR3, as table 1 indicates. The increase of the speed is further benefits from the revolutionary bank-group management technology.

Another differentiating factor is the density, or the physical space of a single memory chip. Developments the memory chip's encapsulation techniques give DDR4 a 50% density increase, with regard to maximum space. This allows DDR4 to pack greater number of bits in the same physical dimensions. However, similar to every emerging new technology, the current price of DDR4 memory is 30% to 50% higher than DDR3, which can achieve similar and comparable functionality. With a large-scale adaptation for DDR4 memory, and hardware compatibility of its peripheral devices, the manufacturing price is expected to reduce to more affordable levels in the future.

GDDR3 and GDDR5, which stand for Graphics Double Data Rate 3 and 5 respectively, are a kind of memory specifically designed for processing graphics. Despite the similarity in terms of the name, the graphic memory is named one generation ahead regular memory. This means that the core technology of GDDR3 is essentially an upgraded version based on DDR2 technology, rather than DDR3. The graphic memory is designed to have lower energy consumption, and an optimized performance when dealing with graphical-data processing. Since the application area of these two kinds memory are fairly different, they do not compete directly. The Graphic DDR is typically developed based on the previous generation of DDR memory technology, with improvements on speed and application-specific functional modifications.

The main markets for our project include traditional memory devices and consumer electronics which are currently in a prosperous state. The global next generation memory technologies market was worth $207.8 million in 2012, and is projected to be worth $2,837.0 million by 2019, growing at a 46.1% average growth rate from 2013 to 2019 (Transparency Market Research, 2014). The report divides the overall market for next

generation memory technologies on the basis of several parameters viz. interface type, application, and geography. On the basis of interface, the market for next generation memory technologies can be categorized into SATA, SAS, DDR, and PCIe and I2C (Transparency Market Research, 2014). The main applications of next generation memory technologies include embedded MCU and smart card, mobile phones, mass storage, cache memory, enterprise storage, and automotive.

Geographically, the global next generation memory technologies' markets can be divided into 4 major categories - North America, Europe, Asia-Pacific, and the rest of the world. This industry is always looking for ways to decrease power consumption, increase density, and develop clever architectures that promote efficient and faster computing. The new generation memory technologies market has gained significant momentum in recent years due to growing demand for faster, highly scalable, and cost-effective memory solutions.

Appreciation for the necessity of our efforts follows an understanding of the dynamic nature of the industry which our product seeks to enter. After understanding the landscape within which we stand, we have good reason to believe that our project is valuable to our stakeholders. We reason that our stakeholders should be more interested in receiving a completed deliverable from us over any other, equally qualified, external competitor.

Our first differentiating quality is that we offer to provide "non-contracted" work. Contracted work is any work commissioned by one party to be executed by another party. To begin such work, both parties must agree on the terms defined within the agreement document prior to the work's commencement. The agreement is realized through means of a binding contract that both parties agree to enact. Once the contract is created, it typically cannot be altered or modified, unless the consent of all parties is evident. This could place the requesting party into a stiff situation if it discovers that its priorities have changed mid-way through a contract.

Upon the project's completion, the completed work is commonly handed off "as-is." This means that no additional support is to be provided in the future (unless explicitly negotiated upon within the original contract). Any additional requested support or

modification requires for a new contract to be written up. Not only is this financially inconvenient, but it can also be logistically inconvenient for the recipient. Without support, the deliverable is handed off with a decreased utility. The recipient of the deliverable is stuck with using the deliverable solely within its original scope.

Our stakeholder, Berkeley Wireless Research Center (BWRC), benefits from ownership over the development process. A common clause added to most contractual work instills a limit on interim design modification requests. Internal control over the developmental process allows for precise design-source malleability during development, and full exposure of the design files. This allows the BWRC professors to more closely guide our direction through the project's development. It allows for them to change the path that we follow if new interests arise. There is no contractual overhead to worry about in this scenario. Owning the source enables the stakeholders to have permanent design-source access without restrictions imposed by third-party trade secrets. Long-term source access enables cost-effective and effort-effective technology adaptation into future projects. Along with adaptation, owning the source creates the opportunity for growing in-house expertise at BWRC through education. These benefits align closely with our stakeholder's interests.

Due to the factors discussed above, our team anticipates BWRC's decision to work with us as opposed to larger suppliers. The current semiconductor marketplace is saturated with both customers and suppliers. "Established operators in this industry have been able to develop solid relationships with customers, and it can be extremely difficult for new companies to gain contracts with customers when existing semiconductor manufacturing operators have built reputations over a long period." (Ulama 2014:28). To exemplify the significance of the previous statement, it is noteworthy that the Semiconductor and Circuit Manufacturing industry is one of the largest exporting industries in the United States (Ulama 2014:5). It indirectly provides jobs to 250,000 Americans, is currently valued at $79.5 billion, and has grown at an annual rate of 4.8% (Ulama 2014:5).

The current players in the industry, both customers and producers, are very well established, and very tightly connected. Penetrating into the customer base that the massive producers currently support is near impossible for a small team like ours due to

lack of reputation and resources. The customers in this segment of the market are a significantly strong force due to two reasons: 1. the intrinsic competitiveness of the current suppliers, and 2. “The electronics marketplace is continually under pressure to improve product functionality, decrease size, increase speed, and decrease cost.” (IBISWorld Global Semiconductor & Electronic Parts 2015:33)

Our team has set our target in a completely different direction to gain competitive advantage. Instead of focusing on the large customers, who are already served very competitively, we direct our focus at an interestingly under-served segment in this market space. In part, our choice of direction is due to the methods through which our Capstone project was decided upon. The decision process confined the scope of the project to target academic goals and provide solutions for academic institutions. Thus, our customer space currently only encompasses the Berkeley Wireless Research Center, but is functionally able to serve any academic or small-scale organizations.

As we currently stand, with one effective customer in our sights, we are subjecting ourselves to a very strong customer market force. This is an undesirable outcome due to the limited size of the space, which we choose to attack, but success in this space will send positive signals to other research institutions. We would be able to expand to encompass more academic institutions because they would prefer to acquire the product through us. Our effective results are comparable to their current methods of operation, but with the benefit of reduced fixed-cost expenditures – which arise when placing orders with large design and manufacture firms.

Majority of the market belongs to well-established companies. These include Micron (IBISWorld 2015: 27) and Texas Instruments (IBISWorld 2015: 30). The limiting factors that lead to this situation include “access to latest technology and intellectual property, the level of investment…, access to skilled employees, and the dominance of existing players” (IBISWorld 2015: 25). In the memory industry, the companies compete over a very specific set of criteria including price, performance, features and power consumption, all of which are highly measurable and quantifiable metrics (IBISWorld 2015: 24).

If the dimensions of competition between companies in a given industry converge, then the companies are left to compete solely on price (Porter 2008: 12). In the integrated

circuits market, the industry has converged heavily on these metrics of performance, features and power consumption, which has resulted in fierce price competition. Because “economies of scale can be significant in this industry” (IBISWorld 2015: 25), new entrants must manufacture large volumes to compete against the bigger companies. This requires up-front capital that many smaller new entrants do not have available. Entering the market attempting to compete on these highly competitive dimensions would result in “zero sum competition” (Porter 2008: 13), and would not be a viable business strategy.

When instead of converging on the same dimensions, companies target different segments of the customer base, the result can be “positive sum” competition: competition that increases the profitability of all companies (Porter 2008: 13). We plan to employ this strategy with our DDR4 memory interface as the needs of a research institution like BWRC are different from the typical semiconductor customer. BWRC focuses on the novelty of the technology and not on the economies of scale. They require the interface to be designed custom to the application of interest and not the general industry standard.

Although the memory technology industry is highly competitive, growing, and difficult to penetrate, the market is growing fast due to this a demand for consumer electronics, an industry which is expected to grow 5.3% annually to nearly $300 billion dollars by 2019 (IBISWorld 2014: 4). This high demand and new market bring some space for new companies to enter and grow. These new entrants usually emerge during the transition between the technological revolutions and each one has its own specialty.

From 2002 to 2013, DDR memory industry has undergone 4 significant technological transitions, all of which are aiming at improving in three performance aspects and achieving a denser data processing capability. Products with “high levels of performance, reliability, quality and low levels of power consumption” (Bach,2014:6) can gain an apparent advance in the competition of memory design industry. Being the three largest manufactures of memory chip and developer of DDR memory technology, Samsung, Mircon and Hynix have already invested millions of dollars in their R&D sector to develop the new generation DDR4 memory interface in order to reinforce their dominating market share.

Given such a giant market, other major memory designers such as Kingston keep fastening their pace to catch up the memory controller design for the recent DDR3 to DDR4 transition. Besides the companies who are already in the market, there are significant number of new entrants, trying to seize this opportunity. The latest Census data indicates that 64.1% of operators in the industry are very small companies and start-ups (Ulama 2014:25). The development strategy of those new entrants are highly focused on certain features, and "specializing in a small number of product lines to serve niche markets" (Ulama 2014:25) in order to avoid a direct competition with large companies.

Table 2 Representative new entrants in DDR4 memory development

Company Name	Specialized market/feature
Century Micro INC.	Small physical size & low energy consumption
Montage Technology	Fast operating speed & low energy consumption
G.SKILL	Enhanced gaming performance

Three unique, representative companies are provided to conduct the analysis of the new entrants. Table 2 above shows a brief comparison of distinguishing factors of the three new entrants. It indicates each new entrant is trying to gain its market share by specializing its product from the three technical aspects mentioned in the previous paragraph.

The Japanese based company Century has just halved the physical size of DDR4 memory in their most recent product at the year of 2014. The China-based Montage Tech is more focusing on developing fast speed and lower power rate DDR4 memory for large scale server use. "Less power draw means less heat and longer battery life", which indicates "the servers are expected to be the biggest beneficiaries of the jump to DDR4" (Andy, 2014:6). Meanwhile, G.SKILL put majority of its resources into developing DDR4 memory controller with improved gaming performance. These companies are

increasingly securing their niche markets by making breakthroughs in design of the memory controller while the major developers are still dominating the memory chip manufacturing area.

Big companies enjoy economies of scale, making it difficult to compete with them in manufacturing the integrated circuits (ICs). Based on the analysis of the new entrants, in order to build immunity for our design, we plan to segment the market to research institutes like BWRC. Their needs are different from most, and provide an opportunity for us to develop a product that satisfies these needs better than the competition. Since the design of our project is specifically for BWRC internal research use, there will be no direct competition and obvious threat from these new entrants either.

The threat from other technologies is weaker, as our DDR4 interface is more advanced than existing DDR3/GDDR5 interfaces. Therefore, we focus on developing the intellectual property and targeting the specific needs of the academic communities. This specific category of consumers require more customizable, and open, circuit designs at a lower volume, a need that is unmet by the larger companies that package their circuits in black boxes, manufacture high volume, and allow little to no customization. By segmenting the market based on unmet needs, and our abilities to satisfy them, we hope to entrench our position as a profitable part of the semiconductor industry.

From the perspective of semiconductor circuit design, it is a complicated process to design an interface and integrate it with the controller and the memory chip. Therefore, our technology suppliers include both software side and hardware suppliers. Software suppliers are those who provide coding languages, design platforms, and simulation tools. Hardware suppliers are those who provide electrical specifications, datasheets, and other fabrication characteristics relating to memory chips.

Software suppliers mainly provide programming language support. Verilog is the main hardware description language that we are using to model digital electronic systems. Cadence, a company that provides electronic design automation (EDA) software, covers many language design platforms, including Verilog and provides EDA tools for full custom design of integrated circuits. As an all-in-one suite, Cadence is our main software supplier.

Hardware suppliers provide descriptive information about the memory chip technology. Each generation of memory chips has new fabrication breakthrough. Thus, during our controller design, the latest information about memory chips is critical, such as voltage supply of the chips and the memory bank structure. Our hardware suppliers, such as Micron Technology, Intel Corp., and Samsung, are big semiconductor companies in this industry. In Semiconductor & Circuit Manufacturing in the US Industry Report, Intel Corp. and Samsung have 18% and 13.8% market share in 2014 (Ulama 2014:4). Although they seem like our competitors from the sales end, they also have the best research departments and technical experts in the chip fabrication domain. Samsung competes in the Semiconductor and Circuit Manufacturing industry via its fabrication and research and development facilities in the United States (Ulama 2014:4). They release academic papers and datasheet of their latest research results about DDR4 memory chip. In accordance with the information provided by these large semiconductor-manufacturing companies, we define the interface and design our memory interface.

Powerful suppliers capture more of the value for themselves by charging higher prices, limiting quality or services, or shifting costs to industry participants. As mentioned above, Intel Corp. and Samsung are both suppliers and competitors for us. If they limit our access to their latest technology about DDR4 memory chip, it will be hard for us to compete with them. However, as the DDR4 memory specification is governed by a standard our dependence on them is reduced.

There are certain aspects that we can focus on to succeed in this capital-intensive, and research-intensive, memory design industry. New companies are trying to explore the market by boosting their expertise in faster-speed designs, smaller dimension layouts, and highly customized application-specific designs. With increasing maturity of the DDR4 technology, the competition is becoming fiercer. This increased competition will largely benefit the semiconductor industry’s evolution speed, as well as provide customers with cheaper and higher efficiency devices. Our project will not only encourage further development from competing companies and research groups, but also benefit BWRC’s exploration of the utilization of DDR4’s capabilities.

3. Intellectual Property Strategy

The Physical Layer (PHY) interface provides us a good scope for creating a patentable Intellectual Property (IP). The physical layer has been split into 5 major parts - serializer, transmitter, receiver, deserializer and timing circuits. Each of these allow for novel implementations and innovations in circuit design. As we are working at the cutting edge of technology, we would have to adopt ingenious techniques to meet the specifications for high data rates of DDR4. One or more of these implementations can provide us patentable IP. This paper will discusses why this technology may be patentable, the advantages and disadvantages of seeking a patent, the current state of the semiconductor IP space, and the risks associated with not seeking a patent.

In the context of IP, creative designs and creative solutions fall under the category of patentable assets. In essence, the purpose for securing IP is to declare discernible ownership over a design or utility (USPTO, 2013). As an independent entity, we can draw benefits from securing patents and owning patents. The benefits we pose to secure range from monetary compensation to strategic industrial presence.

From a monetary perspective, owning patents allows our team to claim ownership to a recognizable asset. After incorporating our team as a legal entity, a patent opens us to the opportunity of being acquired. The proceeds from an acquisition could be used to finance additional ventures, which our team currently does not have the financial freedom to pursue.

A secondary monetization strategy that patent ownership affords us, is the option to license our technology to independent entities who wish to avoid committing R&D expenses for the purpose of developing said technology independently. Aside from the legal expense that we would need to undertake, the licensing option is financially robust.

The third and final benefit is an unquantifiable benefit. It arises from establishing a reputation as an entity. Acquiring a patent will demonstrate that we have the know-how to drive concepts into patentable ideas, and patentable ideas into awarded patents. Successfully acquiring a patent will demonstrate our capability as a team. This reputation will position us to open new leads amongst skeptical and risk averse customers.

The disadvantage of applying patent is obvious: it burns money. Attorney fee is often a big part of the total cost. Determined by the type of invention, the attorney fees range from $5000 to more than $15,000 (Quinn, 2011). Adding the government filing fee and other application fees, the total cost of preparing and filing a patent may exceed ten or twenty thousand dollars. In addition, the maintenance fees would be another big part of the cost. Depending on the number of years the owner intends to keep the patent, the maintenance fees float from $490 for small entities and $980 for large entities due at 3.5 years to $2055 for small entities and $4110 for large entities due at 11.5 years (Stim, 2012).

Considering that this IP would be used only for research or instructional purpose, it would be non-profitable. Therefore, applying for a patent brings financial burden to the owner. In other words, it would not be worth for individuals to applying patent for this IP. However, if the owner switches from individuals to college or Berkeley Wireless Research Center (BWRC), the conclusion would be different. First of all, the college or lab has budget to cover the cost. Furthermore, the patent adds to their reputation, which is far more important than profit for them. Thus, it would be worth to apply a patent for this IP.

Unfortunately, the semiconductor IP market can be difficult for smaller entrants like ourselves. The rate of patent enforcement by larger corporations has not increased over the past few decades (Hall 2007, 5). However, in attempts to increase market share and presence, they have increased the number of patents they file. In the 1980's, the median number of patents filed by an employee was less than one, whereas during the turn of the century it was near eight (Hall 2007, 10).

While larger corporations have a broad and ever expanding portfolio, smaller firms focus on particular market segments in attempts to perfect and own this portion of the total revenue stream. Unfortunately for these smaller firms, this means that if and when larger corporations expand into their territory, they have no choice but to defend the little that they have. It is for this reason that smaller firms tend to more be more aggressive in enforcing their patents (Hall 2007, 3). Thus, it can be expected that we would have to actively enforce and defend our patents.

The risks associated with not patenting the design are significant. Since the design of integrated circuits is based on following certain physical requirement and universal specifications, hundreds of similar design and product can be invented in a short time period based on a same standard. In DDR4 memory design particularly, JEDEC standard is the critical specifications that everyone needs to comply with. There is a high possibility that other individuals or companies will come up with very similar designs. As Gene indicated in his article, engineers who are working on solving a certain problem “are likely to find solutions that are similar” (Gene 2009:8). If a similar design is first patented by other entities, the potential financial loss is irreparable and a great amount design effort would be wasted. Furthermore, without patent the design appropriately, competitors and free-riders can easily take advantage of the design or embedded our inventions into their products without any recognition of our work and having any consequence. Besides these two factors, without right patenting, it is almost impossible to conduct technology transferring or licensing. And this would greatly impede the process of commercialization of the invention or designs.

Therefore, there are a great number of critical risks involving in not patenting the design and novel implementations in our memory interface design should be patented. The management of the patent can be done via creating a patent portfolio. By using management software or having regularly review, updates, categorization and balancing of the patent portfolio, quality management can be achieved.

Trade secret is another kind of intellectual property that has unlimited time of protection. It can also be a method to protect our technology, however it is not the optimal one. Since the DDR technology evolves every three years on an average, unlimited protection time is unnecessary. Giving the fact that circuit design industry is highly standardized and reverse engineering of circuit is quite mature, it would be difficult to protect the design with only trade secrets and no patents since the trade secret sufferers from commercial espionage and high cost of protection. Moreover, the trade secret cannot prevent the similar or same product from being designed. Due the nature of the circuit design industry, trade secret won’t play an effective role in limiting other similar designs. As Shane said, patent is able to “protects your rights regardless of what anyone subsequently develops”

(Shane 2007:8). Therefore, for technological inventions such as circuit design, patenting would be the optimal method to protect its originality.

Ultimately, deciding whether to seek or not to seek a patent for our design depends on the novelty of the final product. If we discover and implement a new physical layer architecture that provides performance, costs, and/or feature improvements over the competition, then the patent's value overcomes the cost associated with filing it. If the final outcome is unique, but provides only marginal benefits compared to the competition, then there will be no benefit in filing the patent.

INDIVIDUAL PAPERS

4. Technical Contributions

4.1 Overview

Our capstone project, ‘Next Generation Memory Interfaces’ aims at the design of high-speed interfaces for memory subsystems. These interfaces allow for data transfer to and from the memory to facilitate interoperability. During the initial phase, we thoroughly studied the unique features and requirements of DDR4, the latest and fastest version of Dynamic Random Access Memory (DRAM) and the implications of those factors for controller and physical layer design.

In this project, my initial focus was on the physical layer that connects the DDR4 memory to its controller. After an extensive literature review, I narrowed my focus to the design of the analog front-end for the receiver on the controller end that accepts the data being transmitted by the memory. The receiver is instrumental in compensating for signal degradation during transmission and ensuring adequate voltage levels at its output to maintain a high signal-to-noise ratio.

The receiver forms an integral part of the communication channel from the memory to the memory controller. It receives the data transmitted by the memory and improves the integrity of the signal for faithful processing by the subsequent blocks in the receive chain. This data is at a higher frequency than the operating frequency of the controller. This is the consequence of serialization, the process of converting multiple parallel streams of data to a single serial stream at a higher data rate than that of the original streams. Serialization is performed at the transmitting end and allows reduction of the routing overhead for multiple transmission lines. However, at the receiving end, the data needs to be deserialized back to parallel streams at the lower frequency that the controller operates at. The receiver also performs one stage of deserialization on the received data, in addition to its fundamental function of ameliorating the signal degradation due to the channel. The output of the receiver is directly connected to the deserializer, a block dedicated to the performing the aforementioned process of deserialization.

4.2 Literature Review

The DDR4 memory achieves twice the data rate of its predecessor, DDR3. This is facilitated by the physical layer, which needs to be capable of reliably handling twice the data rate than it did with the previous version of DRAM, DDR3. At the integrated circuits level, several changes have been made to allow for this. Owing to these changes in specifications, the DDR4 subsystem is capable of achieving the doubled data rates at a lower power consumption allowing for prolonged battery life in devices employing this technology. DDR4 also has increased density which implies greater amount of data storage in the same physical dimensions. All these advantages translate into greater productivity and reduced runtime costs, the two greatly coveted traits in present day technologies.

The physical layer comprises of several circuits and blocks that orchestrate communication and transmission of data between the controller and the memory. It typically comprises of transceivers, SERDES (SERializer – DESerializer) blocks, clock distribution networks, and clock synchronization circuits. Details about the serializer and deserializer are provided in the technical contribution papers of Dillon 2015 and Veryanskiy 2015 respectively. In addition to these it also consists of the sensors and circuits to perform impedance calibration on the data and address buses during operation which aids signal integrity.

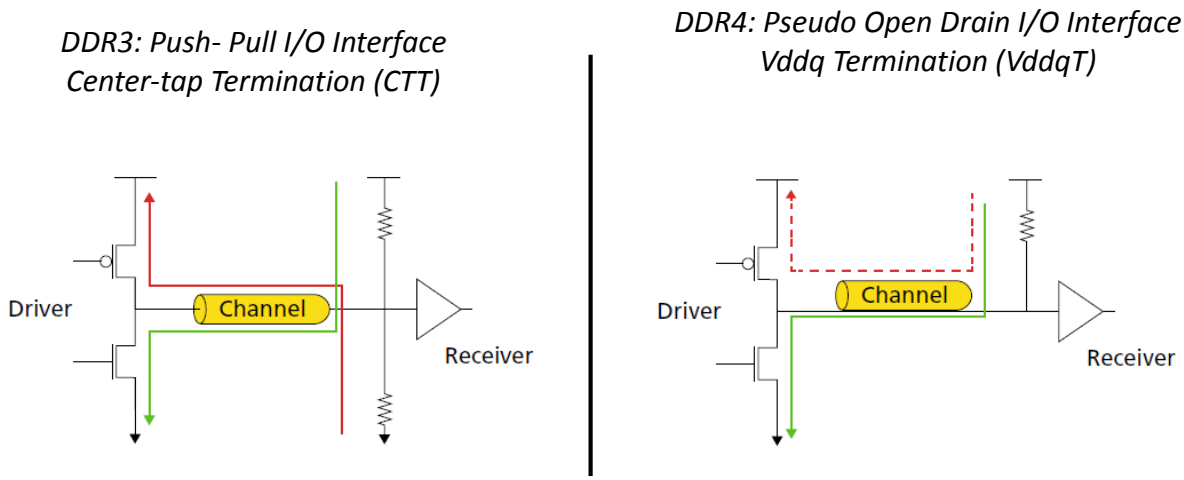

Figure 2 Comparison of the I/O interface and termination scheme in DDR3 and DDR4 (DDR4 Networking Design Guide 2014)

One of the main differences between DDR3 and DDR4 is the I/O interface (Micron Technology Inc. 2014, DDR4 Networking Design Guide). DDR3 used a Push-Pull I/O interface, while DDR4 employs a Pseudo-Open Drain (POD) configuration. The POD configuration allows for low power operation as very little power is required to store a logical ‘one’ (1.2 V) on the line. A pseudo open drain driver has strong pull-down strength but very weak pull-up strength. The term pseudo implies that the pull-up strength is merely weak but not zero, as is the case with the open-drain configuration. In contrast with POD drivers, push-pull drivers have both, strong pull-up as well as strong pull-down networks which consumes more power. The JEDEC Solid State Technology Association defines the standards for the 1.2 V pseudo open drain interface which the controller and physical layer must also adhere to.

DDR4 uses On-Die Termination (ODT) wherein the termination resistors are located inside the chip. The concept is not novel to DDR4 and has been used in the earlier versions of DRAM (Micron Technology Inc. 2008 High-Speed DRAM Controller Design). At the high frequencies of operation of DDR4, the transmission lines in the communication channel cannot be treated as lumped components. Their electrical properties and length have an impact on the signal. One of the major concerns is the signal reflection due to impedance mismatch at the source and load (Feng et al. 2013). To minimize these reflections and their adverse impact on the signal quality, appropriate termination impedances are required. ODT entails that these impedances be present on the die itself, and not connected externally to the chip. The exact value of these impedances depends on the operating conditions such as supply voltage and temperature and therefore, need to be dynamically varied. To achieve this. ZQ calibration is performed. The calibration control block consists of sensing and control circuits that determine the most appropriate value of the termination impedance. ZQ calibration is performed both automatically and upon request by the controller depending on the mode of operation of the memory.

To comply with the pseudo open drain I/O interface, VddqT termination scheme has been adopted in DDR4 as shown in Figure 1. This is another difference between DDR4 and DDR3, as the latter uses a centre-tapped termination scheme. The VddqT termination

compensates for the reduced pull-up drive strength in the POD interface. The changed termination scheme has a direct impact on the receiver front-end circuits which have to now deal with a much wider common mode or reference level range.

The modified termination scheme necessitates analysis and design of an optimum driver circuit. Fundamentally, there are three different termination schemes viz. Type-I (standard termination), Type-II (Switched Termination Pre-emphasis) and Type-III (Constant Termination De-emphasis) (Pham et al. 2010). The standard driver provides no boost to the high frequency components in the signal which, at high data rates such as 3200 Mbps, get attenuated due to channel losses. The pre-emphasis driver overcomes the problem associated with the standard driver but it has its own drawbacks due to the dynamic termination scheme. Due to the switched termination, Type-I driver causes reflections at the termination which deteriorate the channel performance. In view of these arguments, the de-emphasis driver is most suited to DDR4 applications as it provides constant termination and de-emphasizes the non-transition bits that only contain lower frequency components.

Another important characteristic of the DDR4 physical layer is the fly-by topology which helps to improve the signal quality. As data rates increase, signal quality gains paramount importance. In the fly-by topology, when multiple DDR4 chips are connected in a subsystem, they all have their own DQ and DQS lines (Data and Strobe paths). However, for the command, and address lines and the clock and control signals, they are connected in series.

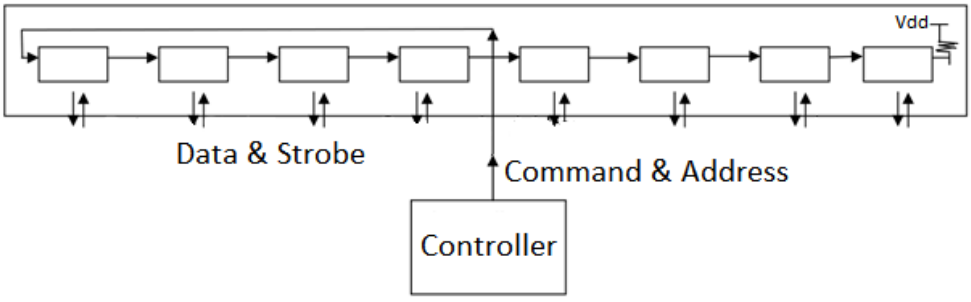

Figure 3 Fly-by topology for Command and Address Lines

The fly-by topology helps to reduce the routing on the board and improves signal integrity by reducing the number of stubs and their length. However, this creates a flight-time skew

between the internal clock and the data strobe signals at every DRAM. To compensate for this skew, write leveling is required. Write leveling adjusts any delay chain and phase settings to align data signals with respect to clock signals for write operations.

To meet the requirements of the high data rate, a robust data fetch scheme is required (Sohn et al. 2013). As discussed earlier in this section, DDR4 uses VddqT termination. While this has a positive impact on reduction of power consumption and parasitic capacitances, it poses a challenge for the receiver. The common mode of the received signal depends on the ODT resistance and the strength of the driver. This implies that the receiver should be capable of operating reliably over a wide input common mode range even in the worst case process, voltage and temperature variations. This issue is resolved through a combination of circuit design techniques which are described below.

The receiver consists of multiple stages, each of which perform a specific, significant task. The first stage receives input signals with a wide range of common-mode levels and converts them to a stable reference level at the second stage inputs by comparing it with a reference (Bucher et al. 2014, Sohn et al. 2013). In general, if the common-mode level of the signal is low, the input device is a PMOS. For high input common-mode levels NMOS devices are used. In this case, as the first stage needs to operate with a wide range of input common mode levels, it consists of an NMOS differential pair, along with a secondary PMOS input pair. Using a combination of the two the receiver can operate over a wide range of common-mode levels. To further ensure that the output common mode level is fixed, feedback is used. Using a technique called common mode feedback (CMFB). CMFB creates a feedback loop that operates only at DC frequency. For the higher frequency, AC signal there is no feedback. Thus, it helps to stabilize the output common-mode level without affecting the signal path in any way.

The second stage is a comparator that helps to pull-up or pull-down the signal to supply and ground levels based on the input signal and reference voltage levels. The comparator can be implemented using current mode logic, allowing high speed operation (Sohn et al. 2013). A drawback of CML comparators, however, is that the output voltage levels are

not supply and ground directly. The CML comparator requires the receiver to have a third stage, a CMOS buffer that converts the low swing signal to appropriate logic levels.

The receiver also addresses the issue of clock domain crossing (CDC) and alignment of the serial data with the strobe signal DQS_EN. The controller and memory operate on different clocks which have the same operating frequency but may differ in phase. This creates a problem for sampling the received data because it is aligned to the transmit clock rather than the sampling clock. To solve this issue a strobe signal is transmitted along with the data which helps to align the edge of the clock with that of the data. This achieved through the use of programmable delay lines and time-to-digital converters (TDC). A TDC creates a digital representation of the time interval between two events (Henzler 2010). Using a TDC, the delay between the receiver clock edge and the strobe signal can be measured. This measurement can be used to adjust the delay of the programmable delay lines to align the data with the clock at the receiving end.

Delay Locked Loops (DLLs) are widely used in memory subsystems for the purpose of clock synchronization and de-skewing. They enable the DRAM to synchronize the output data with the incoming clock signal. DLLs are a critical block which require precision in design and can be energy and area intensive. There exists a tradeoff between accuracy and power in most circuits. In the case of DDR4, accuracy becomes a major concern at the very high data rates as timing margins decrease. However, one of the main features of DDR4 is the reduced power consumption. The challenge, therefore, is to design an accurate DLL with minimum power requirements.

One possible solution to this problem is an adaptive DLL scheme (Sohn et al. 2013). DLLs can be implemented as analog circuits or as digital circuits. Analog DLLs offer greater precision at the cost of higher power consumption. Digital DLLs, although less precise, require lower power for operation. The adaptive DLL scheme seeks to use the advantages of both analog and digital DLLs. It switches between the analog and digital DLL depending on the operating conditions and needs. While this technique helps to reduce power and meets the accuracy requirements, it takes up much larger area to

implement two DLLs while typical systems use only one. It also requires significantly higher design effort to design two DLLs and the integration and control circuits.

Owing to the combination of the techniques described in this section, DDR4 is able to achieve twice the data rate of its predecessor, DDR3 with lower power consumption.

4.3 Methods and Materials:

The first logical step in working in an unfamiliar domain is to develop a strong background in that field through survey and review. In compliance with this idea, an extensive literature review was performed to determine the most critical blocks from a circuits design perspective. We developed an understanding of the specifications and requirements of the DDR4 subsystem and understood the specifications for DDR4 laid down by JEDEC, the committee that establishes global standards for the microelectronics industry. At this stage, my individual focus was on the various components of the DDR4 physical layer and on the circuit design techniques employed in high data-rate systems. Through this I familiarized myself concepts and design techniques widely used in the design of memory subsystems. I also understood the relevance of the differences between DDR4 and DDDR3 from the perspective of signal quality and power consumption. The learning and results from this survey are presented in the literature review section. The first deliverable was to deliver a presentation based on individual surveys to broaden the scope of each team member’s knowledge and to seek guidance from the advisors on the further course of action.

Our second task was to delve further into the circuits that we found most interesting and critical to the system performance. At this step, several research papers about design techniques for the previous versions of DDR memories were read and alternative implementations for the blocks identified previously were investigated. I concentrated on the details of the data fetch scheme and circuit level implementations for the receiver front-end. I paid special attention to the merits and demerits of each circuit topology in the DDR4 context.

The third task involved defining the focus of our project in more depth and creating a block diagram that showed the interfaces between different blocks. This is imperative to ensuring that the blocks designed by different people can fit together seamlessly and that the performance of the system is not compromised at any step. The receiver front-end takes its input from the channel from the memory and provides output to the deserializer. It also performs one step of deserialization. It is important at this stage to create a protocol for the order of deserialization of the data stream received from the memory. The clock frequency and phase used for deserialization also needs to be defined. As the receiver takes the data sent out on the channel by the transmitter, it is of significance to note the range of termination impedances and the best and worst case drive strength of the driver. Each of these are not only relevant from a systems design perspective but also from the viewpoint of designing the internal circuitry of the receiver.

After defining the interfaces, the next step is to define the functionality of each individual block. This is typically achieved through a model that lays down the capabilities and limitations of the block. I used the Verilog-a language to create a model for the receiver front-end. The simplified model comprised of a comparator and an internal reference generator that provides the comparator the reference voltage to compare the incoming signal against. Through a combination of circuits and detailed models it was ensured that all specifications were met and constraints had been accounted for.

Having created a satisfactory model, the next step was to build the circuits for the receiver in accordance with the models. I used Cadence Virtuoso design environment to create the circuits in a 32nm design library. The circuits were tested for performance under variations in process, temperature and supply voltage. They were also tested for accuracy across the entire expected range of inputs and for stability of output. Each of the sub-parts of the physical layer underwent a similar procedure and were tested independent of the other blocks in the system.

Finally, the individually tested blocks were combined and the required adjustments and optimizations were made to ensure that the blocks work well in unison to achieve the

desired functionality. This step was carried out collectively where each person was responsible for adapting their blocks to the system.

The initial literature survey and discussion phase allowed all the members of the team to gain knowledge about the project. This fosters collaboration on the tasks and consistency in the system design. Thereafter, modelling and testing ensures that the various subparts, despite being designed almost independently by different people are compatible and achieve the optimum performance. It also helps to create a standard for each individual block to adhere to. While the modelling and testing phase might appear to be an overhead in terms of time and effort, it is justified the savings in the time required in making the required corrections after the design phase to make the blocks compatible with one another. This procedure adopted in our project allows for smooth operation and seamless integration.

4.4 Results and Discussions:

The analog front-end of the receiver, designed in a 32 nm process, operates at 1.2 V supply and is capable of working reliably at data rates up to 3200 Mbps compliant with the DDR4 standard. The receiver consists of two stages. The first stage deals with the varying common mode level of the received signal and stabilizes it. The second stage, implemented as a Strong-Arm latched comparator, takes this signal and corrects the logic levels in it.

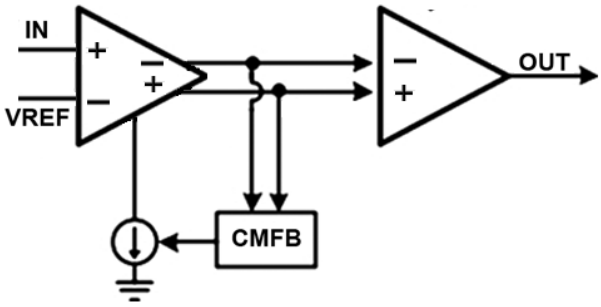

Figure 4 Two stage structure of the receiver front-end

Another sub-block of the receiver front-end is the reference generator. This block provides a reference for the comparator to compare the received data against and determine if it is at logic level 0 or 1. The reference voltage is directly related to the common mode of the received signal. As discussed earlier, owing to the VddqT termination scheme, the common mode of the received signal can vary significantly. Therefore, the reference generator needs to be programmable and capable of generating different levels of

reference voltage. At the same time, it is also required that the generated reference is independent of variations in process, temperature and/or supply voltage.

The receiver front-end also performs one stage of deserialization. This is achieved through the clocked comparator. Two functionally identical comparators are used in parallel. Each of them is clocked on different phases of the clock signal. By doing this, the stream of serial data is converted in two parallel streams, each having half the data rate of the original stream. These two output data channels go to the deserializer for further parallelization.

Pre-Amplifier

The pre-amplifier is the first stage of the receiver front-end. It allows setting the common mode level at the input of the comparator independent of the common mode level of the input signal. The preamplifier provides additional advantages by decreasing the input referred offset of the comparator and the kickback.

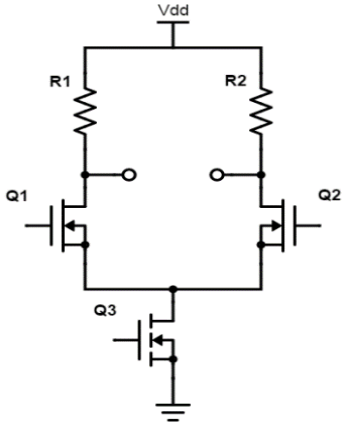

Figure 5 Resistively-loaded Preamplifier

In this technology, a NMOS input amplifier works well across the entire range of input voltages as the output range of the transmitter is limited. As the amplifier does not have a very high gain requirement, it was designed as a resistively loaded, NMOS-input differential amplifier. The resistors on the load side define the DC output voltage level precisely depending on the value of the tail current source. The preamplifier used in the receiver has a gain of about 7 and consumes 20 μA of static power.

Strong-Arm Comparator

As described earlier, the Strong-Arm comparator forms the second stage of the receiver front-end. The first stage shifts the common-mode level of the signal to NMOS input levels. The comparator takes this differential output signal and drives it to VDD or VSS level, depending on the input.

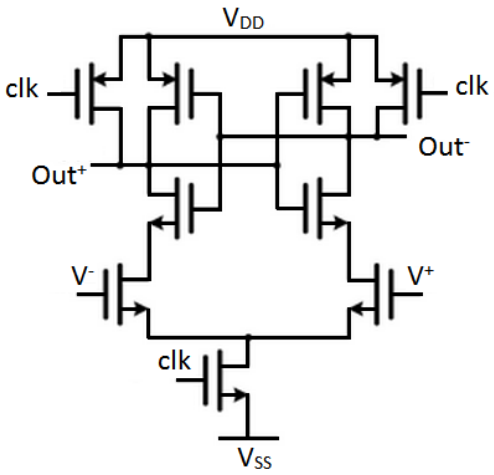

Figure 6 Strong-Arm Comparator

The implementation of the comparator as a Strong-Arm structure rather than a current mode circuit eliminates the need for a third stage because the output levels in a strong arm structure are rail-to-rail, unlike those in the Current Mode Logic (CML) structure. While CML structures have an advantage in terms of their speed of operation, in the present case, in the 32 nm technology even Strong-Arm comparators are capable of meeting the desired specifications.

The Strong-Arm comparator uses a positive feedback latch that allows faster and lower power operation of the comparator. One major advantage of this topology is that there is no static power dissipation which is well suited to the low-power operation of the DDR4 subsystem. The comparator has two phases of operation – reset and evaluation. As the output of the comparator needs to be reset periodically as shown in Figure 6, it is necessary to use latches after the comparator to hold the output at the correct level. The comparator was designed to have a worst case delay less than 312.5ps and minimum power dissipation.

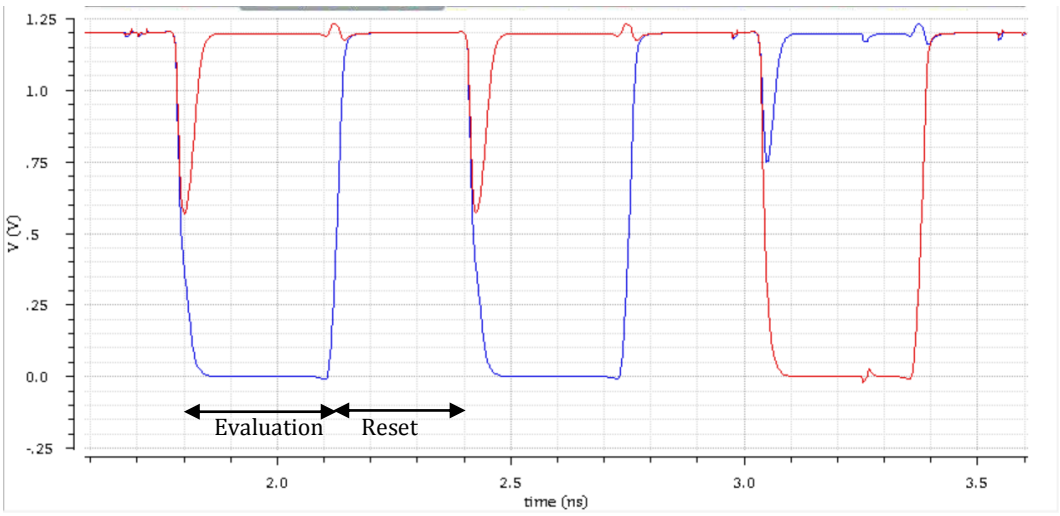

Figure 7 Output waveforms of the Strong-Arm comparator showing the reset and evaluation phases

Integration of the Physical Layer

The physical layer comprises of 4 other blocks – serializer, transmitter, deserializer, and clock generation circuits. Successful operation requires proper interface and communication among these blocks. The interfaces between the blocks are represented in Figure 8. In accordance with this figure, the physical layer was built and tested.

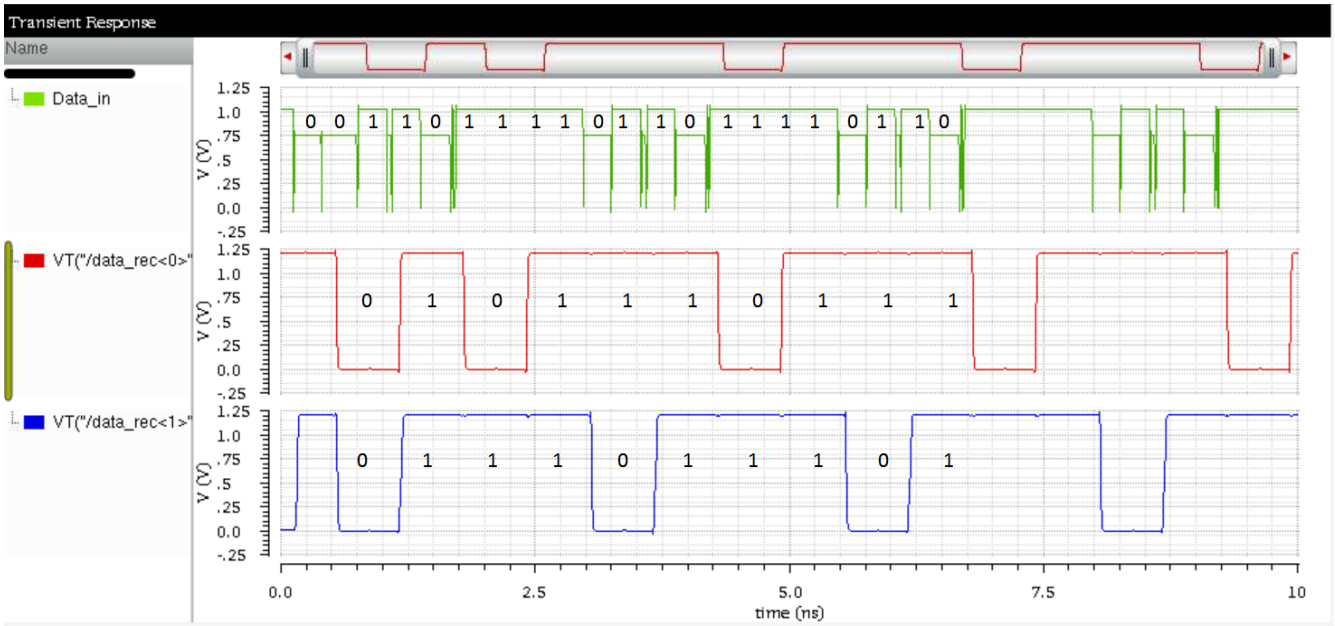

Figure 8 Input and Output data of the receiver front-end showing deserialization

Figure 7 shows the input data (Data_in) and deserialized output data (data_rec<0> [LSB] and data_rec<1> [MSB] of the receiver front-end. It can be easily seen in Figure 7 that the data received (Data_in) gets split into two output streams data_rec<0> and data_rec<1> each of which has twice the period of the original data. The receiver has half a clock cycle of latency before data is available at the outputs.

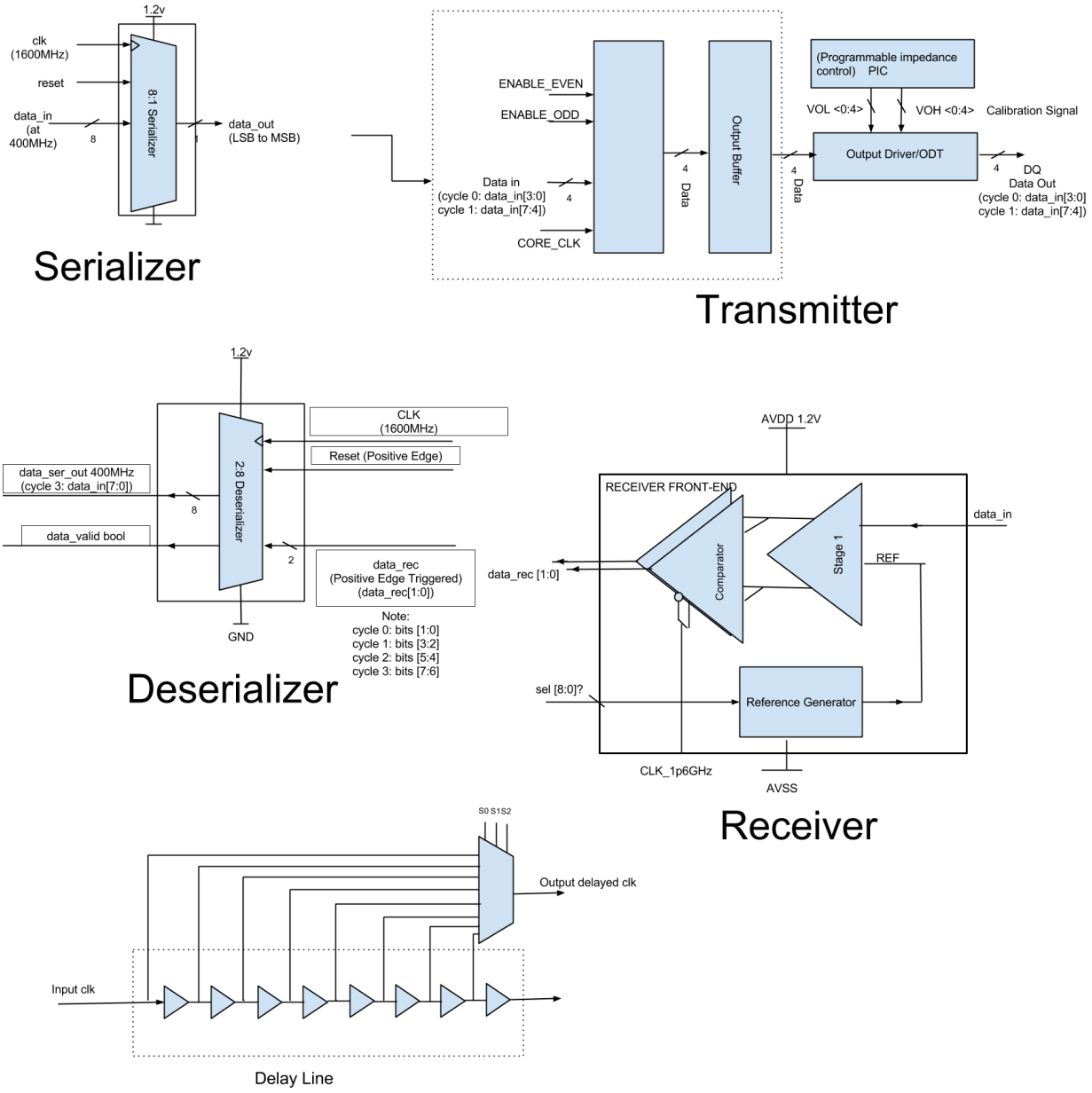

Figure 9 Interface Diagram of the Physical Layer

As discussed earlier, the receiver front-end is an important block in the physical layer. When a signal is transmitted on the channel it suffers from degradation due to the channel imperfections and losses. This effect becomes increasingly significant as the data rate of the channel increases. The receiver helps to counter this degradation by converting the

signal back to the ideal digital voltage levels which makes post-processing easier and more reliable. The receiver takes serial data as input. It is advantageous to transmit data serially as it reduces routing overhead and parasitics. However, the controller requires data as a parallel stream for further processing. Therefore, aside from signal integrity, the receiver also helps to deserialize the data into two streams of half the original data rate. This operation was implemented and verified and the results were presented above.

5. Concluding Reflections

In the project, ‘Next Generation Memory Interfaces’ the physical layer (PHY) interface for a DDR4 memory subsystem was designed in 32 nm Synopsys educational technology. This interface is compliant with the specifications for DDR4 laid down by JEDEC and accounts for all the new features that differentiate DDR4 from its predecessor, DDR3. The interface was divided into five functional blocks that were developed almost independently by the members of our Capstone Project team.

I successfully designed and tested the receiver front-end for the physical layer at the schematic level. Therefore, the initial requirement of design and integration of the physical layer has been met. Throughout the course of the project, a waterfall scheme was adopted for project management. As design of circuits is an iterative process, the scheme was somewhat hybrid during the design and integration phase.

In our project, early identification of risks and potential mitigation strategies has been of paramount efficacy in ensuring timely completion of the tasks. Due to the rather long literature survey phase in the beginning, the project plan and timeline had to be updated a few times. However, this was anticipated and did not cause any significant impact on the final deliverable.

When circuits designed by different people need to be integrated, inconsistencies can cause significant delays and complications. As this situation was expected, we took preemptive measures by extensively defining the interfaces and specifications for each

individual functional block before the design phase. This allowed us to do the integration easily and quickly with minimal changes required to individual blocks.

At its current stage, the project has the schematics for all the five major functional blocks of the physical layer. However, this is only the first step in designing an integrated circuit. The next steps required to realize this circuit to an actual, physical designs would be to do the layout, and post layout verification. The results of this phase might require some changes in the initial design. To convert our physical layer design to a fully functional interface, additional sensing and feedback circuits are required to complete the ODT, ZQ calibration and clock alignment loops. Thus, there is good to scope for further work and research related to our project.

6. Bibliography

1. Bach, Matt Aug. 2014, Tech Primer: DDR4 RAM, http://www.pugetsystems.com/labs/articles/Tech-Primer-DDR4-RAM-589/ accessed March 1st, 2015.
2. Bucher, M., et al. 2014 A 6.4-Gb/s near-ground single-ended transceiver for dual-rank DIMM memory interface systems. IEEE Journal of Solid-State Circuits, 49(1), 127-139.
3. Dillon K., 2015 Next Generation Memory Interfaces Technical Contribution: Serializer.
4. Feng,J., et al. 2013 System level signal and power integrity analysis for 3200Mbps DDR4 interface. Electronic Components and Technology Conference (ECTC), 63rd IEEE 1081:1086.
5. Hall, Bronwyn January 2007 An Empirical Analysis of Patent Litigation in the Semiconductor Industry, accessed March 1st, 2015.
6. Henzler, S. 2010 Time-to-digital converters. Vol. 29. Springer Science & Business Media.
7. Hulkower, Billy August 2013 2013 Laptops and Desktop Computers - US. http://academic.mintel.com/display/637710/ , accessed March 1st, 2015.
8. IBISWorld 2015 IBISWorld Industry Report Global Semiconductor & Electronic Parts. http://clients1.ibisworld.com/reports/gl/industry/default.aspx?entid=960, accessed March 1st, 2015.
9. Micron Technology Inc. 2008, High-Speed DRAM Controller Design, Technical Note TN-04-54
10. Micron Technology Inc. 2014, DDR4 Networking Design Guide, Technical Note TN-40-03
11. Panek, Tracey 2015, Samsung Electronics Co., Ltd. http://subscriber.hoovers.com/H/company360/overview.html?companyId=55277000000000 , accessed March 1st, 2015.
12. Patrizio, Andy 24 Jun. 2014, All about DDR4, the next-gen memory coming soon for PCs and mobile devices. PCWorld,

http://www.pcworld.com/article/2365823/next-gen-memory-is-coming-fast-here-s-what-you-need-to-know-about-ddr4.html, accessed March 1st, 2015.

13. Pham, N., Dreps, D., Mandrekar, R., & Na, N. 2010 Driver Design for DDR4 Memory Subsystems. Institute of Electrical and Electronics Engineers (IEEE).
14. Porter, Michael 2008, The Five Competitive Forces That Shape Strategy, accessed February 16th, 2015.
15. Porter, Michael E. March 1979 How Competitive Forces Shape Strategy. Harvard Business Review https://hbr.org/1979/03/how-competitive-forces-shape-strategy accessed February 16, 2015.
16. Porter, Michael E., and Victor E. Millar. 1985, How Information Gives You Competitive Advantage, accessed March 1st, 2015.
17. Quinn, Gene January 2011, The cost of obtaining a patent in the US, accessed February 28th, 2015
18. Quinn, Gene October 30 2009 The Risk of Not Immediately Filing a Patent Application, accessed March 1st, 2015.
19. Shane, Olafson April 2007 Patents vs. Trade Secrets-Giving Your Business the Competitive Edge, accessed March 1st, 2015.
20. Sohn, K., et al. 2013 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. IEEE Journal of Solid State Circuits 48(1):168-177.
21. Stim, Rich March 2011, The usual costs of filing a patent application, accessed February 28th, 2015
22. Transparency Market Research May 2014, Next Generation Memory Technologies Market (Non-volatile and Volatile) - Global Industry Analysis, Market Size, Growth, Trends and Forecast 2013 - 2019, accessed October 14th, 2014
23. U.S. Patent and Trademark Office October 2013 Description of Patent Types, http://www.uspto.gov/web/offices/ac/ido/oeip/taf/patdesc.htm, accessed March 1st, 2015

24. Ulama, Darryle 2014 IBISWorld Industry Report 33441a: Semiconductor & Circuit Manufacturing in the US. http://www.ibis.com , accessed February 16, 2015.
25. Veryanskiy M., 2015 Next Generation Memory Interfaces Technical Contribution: Deserializer
26. World Intellectual Property Organization May 26 1989 Treaty on Intellectual Property in Respect of Integrated Circuits, http://www.wipo.int/treaties/en/text.jsp?file_id=295136, accessed March 1st, 2015