A 512-Mb DDR3 SDRAM Prototype With $C_{IO}$ Minimization and Self-Calibration Techniques


Abstract—A 1.5-V 512-Mb DDR3 Synchronous DRAM prototype was designed and fabricated in 80-nm technology. Critical to the signal integrity in DDR3 point-to-2points (P22P) interfacing is an efficient calibration scheme and $C_{IO}$ minimization, which were achieved by on-die-termination (ODT)-merged output drivers, SCR type ESD protection, and self-calibration scheme. The hybrid latency control scheme can turn the DLL off in standby mode, reducing power consumption. User-friendly functions such as temperature read-out from on-chip sensor and per-bank-refresh were also implemented.

Index Terms—Calibration, DDR3 SDRAM, input capacitance, per-bank refresh, SCR type ESD, signal integrity, temperature sensor.

I. INTRODUCTION

Main memories have been continuously evolving to be faster, more reliable, and more cost-effective. Although it becomes more difficult to meet all the goals, we achieve a DDR3 SDRAM prototype without losing the sense of balance. Key to the good signal integrity in DDR3’s point-to-2points (P22P) interface operating in the bandwidth from 800 Mbps/pin up to 1.6 Gbps/pin are the minimization of $C_{IO}$ and the competent ODT/OCD calibration schemes, which are realized in this prototype by reorganizing output drivers and calibration schemes. In the 8b-prefetch datapath with three-stage pipelining, a newly devised hybrid-type latency control scheme and a two-step multiplexing can proficiently handle maximum 128-b parallel data in the case of $\times$16 configuration. An efficient protocol for the temperature read-out is proposed, supporting CPU, while it controls the heat in high-speed operations. Per-bank-refresh is another experimental feature of this prototype, virtually removing the loss of the memory bandwidth due to the unavoidable requirement of refresh operation common in all DRAMs. Using triple-metal, single poly-silicon, 80-nm technology, a 512-Mb DDR3 SDRAM was fabricated, showing the performance of DDR1067, preparing the way for the design and production of the fully JEDEC-standard compatible 512M DDR3 SDRAM. The market for DDR3 SDRAM is expected to open in the second half of 2006.

II. ARCHITECTURE

A. I/O Architecture

Fig. 1 shows the chip architecture of 512-Mb DDR3 SDRAM prototype. Memory array comprises eight banks and 8-b prefetch scheme.

![Fig. 1. Chip architecture of 512-Mb DDR3 SDRAM prototype with eight banks and 8-b prefetch scheme.](image-url)
which is an ad hoc way to provide BL4 in 8-b prefetch architecture by discarding the last four data among eight burst data in BL8. To accomplish higher overall bandwidth without increasing that of the DRAM array, DDR3 is designed to prefetch 8-b data instead of 4-b in the case of DDR2. The burst length of 8 (BL8) is, therefore, the innate mode in DDR3. The shorter burst length realized by BL4 chopping, however, can provide better bus efficiency in multi-rank operations, where multiple SDRAMs are accessed in an interleaved way. The timing diagram of BL4 burst chopping is shown in Fig. 3, where the burst length is determined on-the-fly: while read or write command enters, if address 12 is sampled as low, BL4 chopping mode, otherwise, BL8 mode.

B. Latency Control

The clocks from the well-established DLL system [1] and the control signals from the finite state machine (FSM) feed into the hybrid latency control system, whose simplified block diagram is also shown in Fig. 2. The fully parallel latency control scheme [2], although it can provide more bandwidth, does lack the timing margin for turning the DLL off to reduce power consumption in power-saving mode. The hybrid-type latency control scheme can provide reasonable bandwidth without losing the controllability of the DLL. The minimum clock cycle time in the case of the hybrid-type latency scheme is given as

\[ t_{CK}(\text{min}) = \frac{t_{\text{READ}} + t_{\text{SAC}} + t_{\text{uncertain}}}{CL - 0.5 - N} \]  

where \( t_{CK} \) is clock cycle time, \( t_{\text{READ}} \) the time for the FSM to generate the master signal for read operation, \( t_{\text{SAC}} \) the internal DLL clock to data output delay, \( t_{\text{uncertain}} \) the uncertainty time due to process, voltage and temperature (PVT) variation, \( CL \) the CAS latency, and \( N \) the number of shift register(s). The values of \( t_{\text{READ}}(N) \) and \( t_{\text{SAC}} \) can be estimated by HSPICE simulation. Table I provides the calculation results of the maximum bandwidth, in other words, the minimum clock cycle time as a function of \( N \). For values of \( N \) up to two, the hybrid latency control scheme can provide enough bandwidth for DDR3 SDRAM. In this prototype, the value of \( N \) is, therefore, chosen to be one or two as shown in Fig. 2, giving as much margin as one \( t_{CK} \) or two in turning the DLL on. Because the external clock from the CLK buffer, not the internal clock generated by the DLL, triggers the shift register(s), the DLL can have the time to spare for recovery from the standby condition. The hybrid latency control scheme not only meets the bandwidth requirement but also successfully saves more than 10-mA DLL power under the condition of IDD2N (precharge standby current) and IDD3N (active standby current), which is critical in main memory.

III. I/O DESIGN

A. \( C_{IO} \) Minimization—Criteria for Choosing Termination

Since DDR SDRAM, on-die-termination (ODT) has been adopted for better signal integrity, raising the question whether VDDQ, GND, or mid-level of the signal is the best termination. From the standpoint of the potentiality of \( C_{IO} \) minizimization,
most critical to good signal integrity, we have compared two cases: center-tap termination (CTT) [4] and VDDQ termination (VDDQ-T) [5]. Figs. 4 and 5 show that CTT is less sensitive than VDDQ-T to the linearity of driver and to the variation of $R_{ON}$ and $R_{termination}$, supporting CTT as the best choice for good signal integrity in the P22P interface of DDR3 SDRAM.

In the case of VDDQ-T, because the ODT operates in the range of $1.5 \text{ V} \sim 0.6 \text{ V}$, the linearity of driver should be higher than that of the CTT case, where two drivers in CTT—pull up and pull down—divide the operating range by half, reducing the requirement of driver linearity. The better insensitivity of CTT to the linearity of drivers is key to acquiring better signal integrity by the reduction of input capacitance of IO pins.

The output driver comprises the active device (MOS transistor) and passive device (resistor), which are serially connected to each other. The on-resistance of the driver, therefore, comprises two parts: the generally nonlinear active resistance of the pull-up or pull-down driver and the linear passive resistance. The good linearity can be, in general, obtained with larger drivers, which increase both the input capacitance of IO pins and the passive portion of the total on-resistance of the drivers. The total input capacitance to meet the linearity requirement is estimated in Fig. 6.

Fig. 7 shows the estimated jitter in the case of CTT according to the passive portion of on-resistance under various linearity conditions. The transistor $I-V$ curves of output drivers are shown in the upper small figures in Fig. 7. The driver linearity can be defined by the values of the slope when the drain–source voltage ($V_{DS}$) is 0.3 V and 0.75 V. The more negative is the value, the more nonlinear is the driver. The best linearity value is zero. This trend is more critical in the case of the larger passive portion of total on-resistance of drivers (the curve of $\sim 0.1\%$ linearity). In general, input capacitance of IO pins increases with more linear drivers and with larger passive portion of on-resistance, in other words, smaller active portion of on-resistance. The signal jitter deteriorates drastically as the input capacitance of IO pins increases, confirming again that the susceptibility to the problem of a larger driver for better linearity can be a limitation of VDDQ-T in the higher frequency operations of main memory, where $C_{IO}$ can play a dominant role in realizing good signal integrity.

B. ODT-Merged Output Driver and SCR Type ESD Protection

We can learn from Fig. 7 that the $C_{IO}$ minimization is most important in DDR3 and that CTT can survive with less linear drivers to get small jitter, suggesting the $C_{IO}$ can be minimized while evoking as little side effects as possible in the case of CTT. Table II shows the estimated $C_{IO}$ reduction effect mainly due to the ODT-merged output driver scheme. Fig. 8 shows the generic schematic diagram of output driver that is merged with ODT to minimize $C_{IO}$. One-third of the branch would be used during
Fig. 7. Signal jitter estimated according to the passive portion of on-resistance of output drivers under various linearity requirements of output drivers.

Table II

<table>
<thead>
<tr>
<th>Interface</th>
<th>CTT</th>
<th>VDDQ-T</th>
</tr>
</thead>
<tbody>
<tr>
<td>ODT &amp; Driver</td>
<td>Non-linear</td>
<td>Linear</td>
</tr>
<tr>
<td>ODT</td>
<td>0.00 pF</td>
<td>-0.77 pF</td>
</tr>
<tr>
<td>Driver</td>
<td>0.00 pF</td>
<td>+0.25 pF</td>
</tr>
<tr>
<td>ESD</td>
<td>0.00 pF</td>
<td>+0.20 pF</td>
</tr>
<tr>
<td>Total</td>
<td>0.00 pF</td>
<td>-0.32 pF</td>
</tr>
</tbody>
</table>

Fig. 8. Generic schematic diagram of the ODT-merged output drivers.

Fig. 9. Cross-sectional view of the modified SCR-type ESD structure for twin-well technology.

ODT mode and the entire branches during driving mode, which is controlled by ODT command.

By adopting the SCR-type ESD protection, $C_{IO}$ could also be reduced, because the much higher drivability of the SCR device can diminish the driver and ESD junction size. The conventional SCR-type ESD structure, however, needs to be adapted to twin-well technology. We modify the ESD structure as shown in Fig. 9, which can prevent the direct noise coupling between VSSQ and VSS in the twin-well structure, without losing the merits of SCR. Fig. 10 shows the transmission-line pulse (TLP) curve of the modified ESD structure that demonstrates reasonable SCR characteristics. The second breakdown current ($I_{B2}$) is measured to be above 4 amps, which would be enough for the targets of MM (machine model, 100 V) and HBM (human body model, 2 kV) in the case of DRAM.

In summary, by merging ODT into output drivers, around 0.3 pF of $C_{IO}$ reduction is estimated. By adopting SCR-type ESD protection, $C_{IO}$ is expected to reduce by maximum 0.45 pF. From all the experiments, we found that $C_{IO}$ after packaging can be successfully controlled to less than 2.5 pF, which is $C_{IO}$ specification for DDR3 SDRAM. As a result, we can have freedom to decide the appropriate input capacitance of IO pins. The better insensitivity of the CTT interface can be harnessed to make the design of the output driver much easier and more reliable. Although the $C_{IO}$ can be also reduced in the case of VDDQ-T by adopting ODT-merged output drivers as shown in Table II, it is not easy, with the reduced $C_{IO}$, to design output drivers that meet the linearity requirement of the VDDQ-T case.

C. Driver Calibration Scheme

Based on the ODT-merged output drivers with smaller $C_{IO}$ to achieve better signal integrity as well as better reliability, both the newly standardized ZQ-based self-calibration scheme and
the DDR2 SDRAM style external calibration scheme are incorporated, which may be one of the most flexible solutions in memory devices ever reported for ODT/OCD calibration [3]. As shown in Fig. 11, the pull-up current is first self-calibrated by comparing it with the current flowing through the reference resistor connected to a pad newly added in DDR3 SDRAM and named as ZQ, and then the pull-down current is automatically tuned with the pull-up current that is already adjusted. The calibration information is used in updating the registers that generate the signal ODT bits, shown in Fig. 8, to control the value of termination resistance. The driver on-resistance (OCD) is around one-third of that of ODT in the merged output driver.

IV. OTHER FEATURES

A new protocol for temperature read-out is proposed to provide reliable temperature information to the CPU that can be used in cooling down hot memories. To measure temperature in the quietest environment similar to that of temperature sensor calibration during testing the chip, the temperature sensor turns on after EMRS command enters while memory is idle. The 4-b temperature data can be read through pre-defined four DQ pins, before or after the values in the temperature data registers are updated with the response time of \( \sim 5 \mu s \). If the temperature is periodically measured in appropriate time intervals, it is sensible to read out the previously measured data as shown in Fig. 12, preventing the CPU from waiting for \( \sim 5 \mu s \) to get the newly updated temperature data, because the temperature gradient is, for example, known to be less than 1 \( {\degree}C \) in 10 ms. The simplified block diagram for temperature read-out is also shown in Fig. 13.

Per-bank-refresh is cost-effectively realized with little area penalty using one refresh address counter. When all banks are refreshed simultaneously or one by one, the refresh counter is incremented. In case of per-bank-refresh, the counter should be fast enough to successfully increment in \( t_{RRD} \) (bank active to another bank active time) to cope with the case that the two per-bank-refresh operations with different refresh row addresses are sequentially asserted with \( t_{RRD} \) interval. DRAM refresh time can be virtually removed by per-bank-refresh, improving memory bandwidth especially in the case of larger memory density and small page size.

V. MEASUREMENT

The DDR3 SDRAM prototype [6] has been successfully fabricated using triple-metal 80-nm technologies. Fig. 14 shows the...
Fig. 14. Shmoo plot of DQS and DQSB in DDR3 SDRAM prototype.

Fig. 15. Shmoo plot of valid data output window in DDR3 SDRAM prototype.

Shmoo plot of DQS and DQSB, where the minimum cycle time of below 2 ns has been achieved at 1.35 V which corresponds to DDR1067. Fig. 15 shows the shmoo for the valid data output window when $t_{CK}$ is 1.85 ns, where the eye opening is measured about 0.6 UI ($1.0UI = 0.5t_{CK}$). It can be concluded that the designed DLL and data output circuitry successfully operates with a small jitter.

Fig. 16 shows the photograph of the module: raw card A × 8-type unbuffered DIMM, which was assembled with eight DDR3 SDRAM prototype chips. The individual package has dimensions of 11 mm × 14 mm. The UDIMM has, in total, 240 pin-outs: CMD, ADDR, and CLK pins are located at the module center and DQ pins at both sides. Table III summarizes the key features of the 512-Mb DDR3 SDRAM prototype.

VI. CONCLUSION

Most of all, main memories are the general-purpose memories, which should be robust enough to cope with various needs of memory users. They should also be cost-effective under mass production for memory vendors. These two goals are the most vital factors that cannot be sacrificed to achieve better performance such as high speed and low power. The 512M DDR3 SDRAM is one such example. This prototype is designed to be more reliable and cost-effective in achieving the low-voltage and higher bandwidth performance with all the essential key features of the newly standardized DDR3 SDRAM.

REFERENCES


Churoo (Chul-Woo) Park received the B.S. and M.S. degrees in electronics engineering from Seoul National University, Seoul, Korea, in 1990 and 1992, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana-Champaign in 2004.

He joined Samsung Electronics, Korea, in 1992 and participated in the design works for the development of high-bandwidth DRAMs, including SDRAM in 1992, SGRAM in 1994, DDR SDRAM in 1997, and DDR3 SDRAM in 2004—all of which were the first works developed in Samsung Electronics. His research interests are in semiconductor devices and circuit design.
HoeJu Chung received the B.S. and M.S. degrees in electronics engineering from Korea University, Seoul, Korea, in 1997 and 1999, respectively.

He joined Samsung Electronics, Korea, in 1999, where he has been involved in the I/O interface circuit design for high-performance DRAM, such as DDR1, DDR2, Rambus XDR, and DDR3 SDRAM. His current interests are in the area of high-speed digital logic and low-power circuit design for future memory.

Yun-Sang Lee was born in Seoul, Korea, in 1967. He received the B.S. degree in electronic engineering from Sogang University, Seoul, Korea, in 1991.

He joined Samsung Electronics, Korea, in 1991, where he has been engaged in the research and development of next-generation DRAM and SDR/DDR SDRAM. He is currently engaged in the development of low-latency and high-speed memories.

Jaekwon Kim received the B.S., M.S., and Ph.D. degrees in electrical and electronic engineering from the Korea Advanced Institute of Science and Technology (KAIST), Taejon, in 1994, 1996, and 2001, respectively.

He joined Samsung Electronics, Korea, in 2001, where he worked on I/O interface and DLL design for DRAM. He is currently working on protocol and logic implementation on DRAM.

JaeJun Lee received the B.S degrees in mathematics and electronics engineering and the M.S. degree in electronics engineering from Sogang University, Seoul, Korea, in 1997 and 1999, respectively.

He has been with Samsung Electronics, Korea, since 1999, where he has been working as a signal/power integrity Engineer involved in module design for high-speed DRAM interfaces such as RAMBUS, ADT, DRG, and post-DDR3.

Moo-Sung Chae received the B.S. and M.S. degrees in electrical engineering from Seoul National University, Seoul, Korea, in 1998 and 2000, respectively.

In 2000, he joined Samsung Electronics, Korea, where he has been working on the circuit design of high-speed DRAM. He is now with the University of California, Santa Cruz. His research interest is neural recording system and retinal prosthesis.

Dae-Hee Jung was born in Seoul, Korea, in 1971. He received the B.S. degree in electrical engineering from Sungkyunkwan University, Seoul, and the M.S. degree in electronics engineering from Korea University, Seoul, in 1997 and 2004, respectively.

In 1997, he joined Samsung Electronics, Korea, where he has been engaged in the design of high-speed DRAMs such as Rambus and DDR3 DRAM. He is currently involved in the development of high-speed DDR3 DRAM.

Seunghoon Lee received the B.S. degree from POSTECH, Pohang, Korea, and the M.S. degree in electrical engineering from Korea Advanced Institute of Science and Technology (KAIST), Taejon, in 1994 and 1996, respectively.

In 1996, he joined Samsung Electronics, Korea, where he has been involved in the circuit design of SDRAM, Rambus, and XDR DRAM. His research interests are in the design of high-speed memory.

Sung-Ho Choi was born in Pohang, Korea, in 1972. He received the B.S. degree in electrical engineering from Dankook University, Seoul, Korea, in 1997.

In 1997, he joined Samsung Electronics, Korea, where he has been engaged in the design of DRAM.

Seung-young Seo was born in Kunsan, Korea, in 1973. He received the B.S. degree in electronic engineering from Semyung University, Korea, and the M.S. degree from Han Yang University, Korea, in 1999 and 2001, respectively.

He joined the Memory Division, Samsung Electronics, Korea, in 2001, where he has been engaged in the design of DRAM.

Taek-Seon Park was born in Seoul, Korea, in 1974. He received the B.S. degree in electrical engineering from Semyung University, Korea, and the M.S. degree in electrical engineering from Hanyang University, Korea, in 2001 and 2003, respectively.

In 2003, he joined Samsung Electronics, Korea, where he has been engaged in the design of DRAM.

Jun-Ho Shin was born in Masan, Korea, in 1973. He received the B.S. and M.S. degrees in electronic communication engineering from Hanyang University, Seoul, Korea, in 1996 and 1998, respectively.

In 1999, he joined Samsung Electronics, Korea, where he has been engaged in the design of DRAM.

Jin-Hyung Cho was born in Seoul, Korea, in 1978. He received the B.S. degree in electrical engineering from Hongik University, Seoul, in 2001.

In 2001, he joined Samsung Electronics, Korea, where he has been engaged in the design of DRAM.
Ki-Whan Song received the B.S., M.S., and Ph.D. degrees in electrical engineering from Seoul National University, Seoul, Korea, in 1994, 1996, and 2005, respectively.

From 1996 to 2001, he was with the DRAM design team, Samsung Electronics, Korea, where he contributed to the development of RAMBUS DRAM. In 2005, he rejoined the Advanced Technology Development team, Samsung Electronics, where he has been working in the area of high-density memory design. His research interests include silicon nano-electronic devices, SONOS nonvolatile memories, and one-transistor DRAM.

Kyu-hyoun Kim was born in Seoul, Korea, in 1972. He received the B.S., M.S., and Ph.D. degrees in electrical and electronic engineering from the Korea Advanced Institute of Science and Technology, Taejon, in 1992, 1994, and 1998, respectively.

He joined Samsung Electronics, Korea, in 1998 and has been developing DDR, DDR2, DDR3, and high-speed DRAM. He is the leader of the high-speed I/O circuit development group for future DRAM products.

Jung-Bae Lee was born in Seoul, Korea, in 1967. He received the B.S., M.S., and Ph.D. degrees in electronics engineering from Seoul National University in 1989, 1991, and 1995, respectively.

In 1995, he joined Samsung Electronics Company, Ltd., Gyeonggi-do, Korea, as a Research Engineer, where he has been working on the circuit design of high-speed DRAM. His research interests include design of high-speed memory and VLSI systems and noise phenomena in electronic systems and semiconductor devices.

Changhyun Kim (SM’05) received the B.S. and M.S. degrees in electronics engineering from Seoul National University, Seoul, Korea, in 1982 and 1984, respectively, and the Ph.D. degree in electrical engineering from the University of Michigan, Ann Arbor, in 1994.

In 1984, he joined Samsung Electronics Company, Ltd., Gyeonggi-do, Korea, where he has been involved in circuit design for high-speed dynamic RAMs, ranging from 64 Kb to 16 Mb in size. From 1989 to 1995, he was a Research Assistant and Research Faculty with the Center for Integrated Sensors and Circuits, University of Michigan. In 1995, he rejoined Samsung Electronics and worked on Gigabit DRAM and high-speed devices, such as RDRAM and DDR2. His present research interests are in the area of circuit design for low-voltage and high-performance Gigascale DRAMs and future high-performance memory architectures (DDR3, XDR, NMT, etc.), ranging from 1 GHz to 10 GHz in speed. Since 2005, he has been in charge of the Advanced Technology Development team for the future new memory designs including DRAM, SRAM, Flash, and other emerging new memories.

Dr. Kim was selected as a Samsung Fellow in 2004 and received several Grand Prizes from the Samsung group for the successful development of 1 Mb, 1 Gb DRAM, and 144M RDRAM in 1986, 1996, and 2000, respectively. He received several technical achievement awards from the R&D Center of Samsung for his work on the development of high-speed devices and the characterization of submicron devices and reliability issues in high-density DRAMs, including reducing soft-error rates and reducing sensitivity to electrostatic discharge problems. He received First Prize for Design Excellence in student VLSI design contests at the Center for Integrated Sensors and Circuits in 1991 and 1993, sponsored by several U.S. companies. He serves as a committee member of the Symposium on VLSI Circuits and Asian Solid State Circuits Conference.

Soo-In Cho was born in Seoul, Korea, in 1957. He received the B.S. degree in electronics engineering from Seoul National University, Seoul, in 1979 and the M.B.A. degree from Ajou University, Suwon, Korea, in 2001.

He joined the Semiconductor Research and Development Center, Samsung Electronics, Gyeonggi-do, Korea, in 1979, where he was engaged in the design of CMOS logic LSI. Since 1983, he has been in charge of circuit design for high-speed DRAMs. He is currently the Chief of the Department of DRAM Design, Samsung Electronics.