A 0.83-pJ/bit 6.4-Gb/s HBM Base Die Receiver using a 45° Strobe Phase for Energy-Efficient Skew Compensation

Sangyoon Lee, Han-Gon Ko, Student Member, IEEE, Joo-Hyung Chae, Member, IEEE, Soyeong Shin, Student Member, IEEE, Jaekwang Yun, Deog-Kyoon Jeong, Fellow, IEEE, and Suhwan Kim, Senior Member, IEEE

Abstract—Skews between data and strobe signals can occur in HBM transceivers due to process and voltage variations across the base die. Skew compensation is introduced into the deserializers of our quarter-rate single-ended receiver for next-generation unmatched source-synchronous HBM interfaces. Data and strobe signals are energy-efficiently realigned by using a 45° strobe phase DQSϕ. This phase, which is equidistant between quadrature strobe phases DQSϕ and DQSϕ0, is generated by a digital type phase interpolator of our receiver. The transceiver, including the proposed receiver, was designed and fabricated in a 65nm CMOS process. The receiver corrects skews to within 7.8ps at a data-rate of 6.4Gb/s, with an energy cost of 0.83pJ/bit per pin.

Index Terms—Quarter-rate receiver, single-ended signaling, memory interface, unmatched source-synchronous scheme, skew compensation scheme, HBM base die, phase interpolation.

I. INTRODUCTION

It is becoming difficult for standard DIMM to meet the need for increasing data-bandwidth because this type of memory is effectively limited to 50GB/s by pin count and signal integrity issues [1]. The bandwidth of memory can be increased by reorganizing its internal layout, with the introduction of stacked DRAM cells, through-silicon vias (TSVs) and a large number of silicon interposer I/O channels: this is referred to as high-bandwidth memory (HBM) [2]. An HBM2 chip has a large number of pins, which allows it to operate at a bandwidth of 256GB/s with single-ended signaling [3]. The next generation of HBM, HBM3, is expected to have a bandwidth beyond 512GB/s, implying a data-rate of at least 4Gb/s per pin with single-ended signaling.

As an HBM base die has to be connected to all the micro-bumps and TSVs in the chip, it is large enough to experience skews between data (DQ) and strobe (DQS) signals due to local variations [4]. The effect of skew appears even severer in high-speed interfaces because reduced setup and hold-time margin can cause timing violation [5]. One way of coping with increasing skews while maintaining the power budget is to re-train the DRAM at intervals [6], but this requires time and space. A low-power skew compensation offers a more promising alternative.

Skew compensation [7, 8] can be performed in memory interfaces by phase-shifting the data signals to their optimal sampling-point. This is adequate to compensate for the skews between data and strobe signals in matched source-synchronous schemes, in which data signals are buffered to CMOS level before being sampled, as shown in Fig. 1(a). However, at data-rates over 4Gb/s, unmatched source-synchronous schemes [9], in which data signals are directly sampled without being buffered to CMOS level have been more generally adopted in DRAM interfaces. At the expense of unmatched delay between DQ and DQS, this arrangement shown in Fig. 1(b) can significantly reduce power as data buffers are omitted. It is even more preferred in HBM interface where there are a large number of DQ channels.

Forwarded-clock clock and data recovery (CDR) circuits based on injection-locked oscillators (ILOs) [10] are capable of compensating for skews in unmatched source-synchronous schemes without much power consumption. However, because ILOs require a clock source that is always running, this type of CDR is not applicable for DRAM interfaces, which use strobe signals to sample DQ signals and have idle states. A digital DLL-based clock generator [11] can be used for skew compensation in unmatched source-synchronous memory systems. However, the generation by the DLL of extra clock edges for oversampling uses too much power to be generally acceptable [12] in HBM interfaces. Moreover, both of these schemes use differential signaling, which is incompatible with DRAM interfaces, which are usually based on single-ended signaling to reduce the pin count.

To overcome the drawbacks described above, we propose a single-ended energy-efficient quarter-rate skew-compensated receiver. It incorporates skew compensation schemes appropriate for unmatched source-synchronous memory.
Fig. 2. Architecture of the proposed skew-compensated receiver.

Fig. 3. (a) Block diagram and (b) timing diagram of the proposed skew-compensated deserializer.

interfaces. The receiver consists of a skew-compensated deserializer and a DQS path that has a digital type phase interpolator. The interpolator, generating a single-fixed-phase of 45°, does not require much burden in terms of the receiver’s power consumption per pin. The 45° strobe phase is used to oversample for skew compensation. To validate the practicality of the proposed receiver, an entire transceiver, including an eye monitoring circuit and emulated on-chip silicon interposers, was designed and implemented in 65nm CMOS technology.

The rest of this paper is organized as followed: in Section II the proposed receiver is described; in Section III the transceiver for testing the receiver is introduced; in Section IV experimental results are presented, and in Section V conclusions are drawn.

II. RECEIVER ARCHITECTURE

The architecture of the proposed skew-compensated deserializer is shown in Fig. 2. The receiver has an unmatched source-synchronous scheme appropriate for HBM interfaces, in which the samplers directly sample the DQ signals that are transmitted over the channel without buffering them to CMOS swing level. Therefore, a quarter-rate clocking is used to provide regeneration time for the samplers. The receiver consists of a DQS path with a 45° phase interpolator and skew-compensated deserializers. The differential current mode logic (CML) signals DQS_T and DQS_c are buffered to CMOS swing level by the clock buffer (CLKBUF) in the DQS path. The outputs of this are divided into 4-phase clock signals by the IQ divider (IQDIV) to provide quadrature strobe signals. To align the rising edges of the quadrature DQS phases with the center of each unit interval of DQ, an additional phase of the strobe signal that is equidistant between 0° and 90° strobe phase is introduced. This phase is interpolated by the 45° phase interpolator (45° PI), which uses two consecutive phases of clock to generate a mid-phase clock. The 45° PI is located at the end of the DQS path, where the quadrature strobe signals are buffered. Unlike phase interpolators that generate a number of phases [13], our digital type phase interpolator was implemented to generate a fixed strobe phase (DQS_{45}) to minimize power burden for skew compensation.

A. Skew-Compensated Deserializer

Fig. 3(a) shows the block diagram of the proposed skew-compensated deserializer (SCDES). The DQS path provides the SCDES with the quadrature DQS signals DQS_0, DQS_90, DQS_{180}, and DQS_{270}, and DQS_{45}, which is equidistant from DQS_0 and DQS_90. The SCDES uses DQS_{45} to sample DQ, generating the signal SAMP, which is sent to a digital loop filter. This filter, which also receives the deserialized outputs, returns a series of digital control words (DCWs), which are used to change the delay of digitally controlled delay lines (DCDLs) to compensate for any lead or lag in the DQS signals, and to center them on the eye of the DQ signal, which is the optimal sampling point. For as long as there are transitions in DQ, the SCDES compensates for the skews between the data and strobe signals. The DCDLs, which are made of shunt-capacitor inverter delay cells [14], are controlled by six binary bits and have a resolution of 5.1ps/code at a receiver supply voltage $V_{DDRX}$ of 1.1V.

The timing diagram shown in Fig. 3(b) illustrates the operation of the SCDES. In Case 1, DQS_{90} is ahead of its optimal sampling-point, and therefore, the sampling margin is inadequate. At this juncture, the value of SAMP equals that of OUT_0. When the transition of DQ is detected, the values of OUT_0 and OUT_90 become different, and the digital loop filter returns a DCW with a larger value, which moves the DQS edges towards realignment. In Case 2, the rising edges of DQS lag their optimal sampling-point, and the value of SAMP equals that of OUT_{90}. When a transition in DQ is detected, the digital loop filter returns a smaller DCW. When no transitions in DQ are detected, the values of OUT_0 and OUT_90 are the same, the digital loop filter does not operate, and so the DCW value is unchanged. This arrangement saves power during the idle periods encountered in memory interfaces.
 interpolate NODE \( Q \) digital loop filter until it becomes close enough to DQS \( I \). Note that the delay of CDL1 is made twice that of \( \Delta t \) and \( 0.5(\Delta t+1UI) \) from the rising edge of DQS \( I \), respectively. \( \Delta t+0.5UI \) and \( \Delta t+1UI \). Thus, the rising edge of NODE\( M \) is shifted by \( \Delta t \) and \( \Delta t+0.5UI \) from the rising edge of DQS \( I \). If NODE\( M \) is shifted by \( \Delta t \) to generate DQS\( 45 \), and DQS\( 90 \) is shifted by \( \Delta t \) to generate DQS\( 90, \) then the difference between the phases of DQS\( 0 \) and DQS\( 90 \) will stay the same as that between the phases of DQS\( I \) and DQS\( 90 \), while the signal DQS\( 45 \) is generated, with a phase mid-way between them. The rising edges of DQS\( 0 \), DQS\( 45 \), and DQS\( 90 \) are \( \Delta t \), \( \Delta t+0.5UI \), and \( \Delta t+1UI \) apart from the rising edge of DQS\( I \). Fig. 4(b) shows the implementation of the proposed 45° PI. It consists of coarse delay lines (CDL1 and CDL2), which are made of NAND delay elements [15], inverters, and a digital loop filter. Note that the delay of CDL1 is made twice that of CDL2 by introducing a replica delay line. The digital loop filter compares the phases of NODE\( I \) and DQS\( 0 \) and changes the coarse delay line control word CCW until their phases are sufficiently close. The timing diagram of the 45° PI is shown as Fig. 4(c). When the 45° PI powers up, NODE\( I \) is shifted by the CCW from the digital loop filter until it becomes close enough to DQS\( 0 \) to interpolate NODE\( M \). Note that inverter delays are neglected in the timing diagram. When the rising edges of NODE\( I \) and DQS\( 0 \) are far apart, the phase of NODE\( M \) is undefined; it is by no means certain that the phase of DQS\( 0 \) is equidistant from those of DQS\( 0 \) and DQS\( 90 \). Once the digital loop filter judges that NODE\( I \) and DQS\( 0 \) are sufficiently close, CCW is locked and NODE\( M \) is interpolated and shifted to generate DQS\( 45 \). The addition of the 45° PI increases the DQS path delay \( (TDQS2DQ) \) by \( \Delta t \). However, the delay variation due to corner change in CLKBUFF of the DQS path is the most dominant. Also, as the PI delays \( \Delta t \) by setting 1UI as the target through calibration, the addition of the 45° PI to the DQS path does not vary \( TDQS2DQ \) significantly due to corner changes. A corner simulation result of the 45° PI is shown in Fig. 4(d). It shows rising edge times of DQS\( 0 \), DQS\( 45 \), and DQS\( 90 \) relative to that of DQS\( I \) in various corners written in the figure. In all the combinations of the corners, the 45° PI successfully generates the mid-phase of DQS\( 0 \) and DQS\( 90 \), with a variation of less than 5% of a unit interval from the desired signal at a data-rate of 6.4Gb/s. The DQS path, including the 45° PI, which generates the quadrature phases and the extra interpolated phase of DQS, distributes these phases to different DQ pins. In an HBM base die, 32 DQs share one DQS\( T \) and DQS\( C \) pair; thus, the extra power consumption by the 45° PI is not burdensome when considered per pin. III. TEST STRUCTURE As shown in Fig. 5, we designed the transceiver with emulated silicon interposer channels to verify the proposed receiver. A pattern generator receives an external clock signal CLKE\( X \), from which it generates the signals D0 to Dn, and STB, which are then serialized by 8:1 serializers. As well as generating continuous signals, the transmitter can produce the bursts of signals typically experienced by a DRAM interface. To measure the bit error-rate (BER) of the output, the serialized signals go through the digitally controlled delay lines in the transmitter (TXCDCLs), and are then sent to emulated silicon interposer channels by N-over-N drivers (NNDRV) which operate with a supply voltage \( V_{DDQ} \) of 0.4V. After transmission over the channels, the signals DQS\( I \) and DQS\( C \) arrive at the DQS path, where they are buffered to a CMOS level, and quadrature phases are generated, together with the extra phase DQS\( 45 \). The SCDESs use these signals to sample the DQ signals in the process of compensating for skews and deserialization. An on-chip eye monitoring circuit also receives the DQ signals and uses an external clock CLKE\( Y \) to analyze the extent to which they are affected by transmission over the channels. It also analyzes the resolution of TXCDCL and the offset caused by the layout mismatch of the comparators. An output multiplexer selects the signal to be analyzed, and a driver transmits it. IV. MEASUREMENT RESULTS A prototype chip was fabricated in a 65nm CMOS technology. Fig. 6 shows the die micrograph and the measurement setup. The clock signal CLKE\( X \) is supplied by the pattern generator of a J-BERT (Agilent N4903A). To see whether DQS\( 45 \) is correctly interpolated between DQS\( 0 \) and DQS\( 90 \), the DQS signals were displayed on an oscilloscope.
Direct off-chip monitoring of a signal after transmission over the channel cannot be achieved without affecting that signal, and so an eye monitoring circuit was implemented to see whether the signal is transmitted satisfactorily. Fig. 7(a) summarizes the algorithm that this circuit uses to measure the eye of a DQ. The BER of DQ is sampled by CLK\(_EYE\) while sweeping the TXDCDL code so that the data is delayed before being transmitted to the channel by NNDRV. Each BER is measured by the J-BERT and sent to a PC, which then increments the TXDCDL code through the I\(_2\)C interface. The BERs for every value of TXDCDL are measured while changing the value of reference voltage VREF in on-chip eye monitoring circuit, this producing the eye diagram of the signal.

Fig. 7(b) shows how the BER of the deserialized outputs of DQs are measured by the J-BERT. First, VREF is fixed to 200mV, which is half of V\(_{DDQ}\), and then the TXDCDL codes of DQS are swept by the PC through the I\(_2\)C interface. Changes in receiver supply voltage V\(_{DDRX}\) shift the sampling-point of DQS, as shown in Fig 8. The BERs measured at different values of V\(_{DDRX}\) show how much DQS is shifted by receiver supply voltage variation. The bathtub curves in Fig. 8 show how the RX voltage affects the extent to which the rising edges of DQS signals are skewed. From these curves, we can determine the DQS TXDCDL code that corresponds to the center of the bathtub curve measured at the design voltage of 1.1V. This code, which is 80, is used to initialize the SCDES, for any value of V\(_{DDRX}\) before skew compensation. The value of the DQS TXDCDL code is modified during the skew compensation process, and this final value is used for the subsequent BER measurement, as explained in Fig. 7(b). If the skew is correctly compensated, bathtub curves like those shown in Fig. 8 should all be centered at the DQS TXDCDL code of 80, irrespective of the value of V\(_{DDRX}\).

Fig. 9 shows the output waveform from the 45° PI, measured using an oscilloscope. At a data-rate of 6.4Gb/s, the rising edges of DQS\(_0\) and DQS\(_90\) are one unit interval apart, where the unit interval is 156.25ps. Within the accuracy of the oscilloscope trace, the rising edge of DQS\(_{45}\) appears to be equidistant from the rising edges of DQS\(_0\) and DQS\(_90\), with half a UI between each successive rising edge. The power consumption of the DQS path was found to be 5.7mW at a data-rate of 6.4Gb/s and a supply voltage V\(_{DDQS}\) of 1.1V, of which 2.2mW was consumed by the 45° PI.

Fig. 10 shows eye diagrams of DQ0 and the resolutions for two sample chips, measured at a data-rate of 6.4Gb/s with a PRBS7 data pattern. Since the VREFGEN was designed with resistor ladders, its resolution was calculated from the ratio of the designed resistances. To find the resolution of TXDCDL, the middle codes at VREF of the largest horizontal opening were measured for two adjacent eyes. Then, the difference between the two middle codes was divided by the data-rate to calculate the TXDCDL resolution.
Fig. 11. Measured bathtub curves showing the variation in receiver voltage (a) before and (b) after compensation.

Fig. 12. Measured skews at a selection of three DQ pins from each of two sample chips, for extreme receiver supply voltages of (a) 1.0V and (b) 1.2V. The black points are the skews before compensation and the red points are the skews after compensation by the SCDESs.

V. CONCLUSIONS

Table 1 summarizes the performance of the proposed receiver and compares it with that of other receivers capable of compensating for skews in source-synchronous memory interfaces. Our receiver has a quarter-rate architecture and uses a single-ended unmatched source-synchronous clocking for compatibility with next-generation HBM interfaces. It has the best FoM of those compared in Table I, which is 0.83pJ/bit per pin at a data-rate of 6.4Gb/s. Because of the simple structure, the digital type phase interpolator to generate a fixed 45° phase consumes low power. Also, using only one additional phase to oversample requires minimum power burden for skew compensation. The FoM should be even lower when the power consumption of the DQS path is amortized across 32 DQ pins in a real HBM base die [3]. A single SCDES takes up an area of 0.0051mm², while the DQS path occupies 0.0064mm².

REFERENCES