

## 6.5 A 64Gb/s PAM-4 Transceiver Utilizing an Adaptive Threshold ADC in 16nm FinFET

Luke Wang<sup>1</sup>, Yingying Fu<sup>2</sup>, MarcAndre LaCroix<sup>3</sup>, Euhan Chong<sup>3</sup>, Anthony Chan Carusone<sup>1</sup>

<sup>1</sup>University of Toronto, Toronto, Canada

<sup>2</sup>Huawei, Markham, Canada

<sup>3</sup>Huawei, Ottawa, Canada

ADC-based transceivers having up to 8 bits of resolution have been reported for PAM-4 links above 50Gb/s [1,2], although fewer bits are sufficient and offer lower power for short reach (SR) channels. To further reduce the power consumption of ADC-based wireline transceivers, non-uniform quantization has been explored [3,4] using performance metrics for the complete link, such as bit-error rate (BER), to optimize the quantizer thresholds. Both [3,4] are PAM-2 (NRZ) receivers, demonstrating non-uniform quantization specifically for a decision feedback equalizer (DFE) at 10Gb/s and a feedforward equalizer (FFE) at 4Gb/s respectively. An LMS algorithm in [4] adjusts the threshold levels requiring fine-tuning (8b resolution). This paper presents a 64Gb/s PAM-4 transceiver utilizing an ADC-based receiver (RX), with an analog front-end (AFE) based on a 6b, 1b folding, flash ADC with adaptive threshold levels. A fast greedy-search algorithm is used to choose the optimal quantizer thresholds for minimum BER over a given channel. This provides a near-optimal way of power-scaling the ADC when the channel loss doesn't require the ADC's full resolution. The optimization can work in the background for any equalizer structure, does not place additional requirements on the ADC design, and never diverges, unlike LMS-based approaches [4].

The transmitter, Fig. 6.5.1, consists of separate 32:4 serializers for the MSB and LSB data streams and 3 identical driver clusters. The MSB serializer drives 2 driver clusters while the LSB serializer drives 1 cluster. In NRZ mode, identical data streams are applied to both MSB and LSB serializers. Each driver cluster comprises 11 SST slices: 6 dedicated to the main pre-emphasis tap only, 2.5 driven by either the main or pre-cursor tap, and 2.5 driven by either the main or post-cursor tap. The half slices are needed to increase the pre-emphasis tap weight resolution in PAM-4 mode. The slice-based design enables termination calibration and facilitates nonlinearity compensation: relative weights of MSB and LSB slices can be tuned to adjust the center eye in steps of  $36mV_{ppd}$ , and the upper and lower eyes in steps of  $18mV_{ppd}$ . This allows the RLM to change from 0.99 to 0.89 if an asymmetrical eye is desired. A CML stage in parallel with the SST driver slices increases output voltage swing by a further  $100mV_{ppd}$  (10%), similar to [5]. The TX uses quarter-rate (QR) clocks for the 4:2 multiplexer (MUX) and a half-rate (HR) clock for the final 2:1 MUX. The HR clock is created by XORing the quadrature QR clocks. The skew between the QR and HR clocks is minimized by design and verified in post-layout simulations across corners. A duty cycle correction (DCC) circuit provides fine adjustment with 250fs resolution by controlling delay on the rising edge of *clock* and falling edge of *clockb* as shown.

The RX AFE, shown in Fig. 6.5.2 consists of an 8-way time-interleaved 32GS/s 6b flash ADC with 1b folding. The front-end sampler based on [6] is fitted with degeneration to provide  $\sim 6dB$  boost at  $f_s/2$ . It is driven by an HR clock and acts as a 2-way time-interleaver. Hence, as long as accurate DCC is performed on the HR clock, up to 1UI skew is tolerable on the subsequent  $f_s/8$  sub-ADC clocks. The duty-cycle corrector is designed with a range of  $\pm 10\%$  and a resolution of approximately 0.35%. Each sampler output directly drives 4 sub-ADCs. Each sub-ADC has a PMOS track-and-hold (T&H) at its input, which is in track mode for  $< 2UI$  per conversion to ensure that only 1 T&H loads the sampler (on either side) at a time. The BW of the entire sampling network including interconnect, at the minimum CTLE boost and maximum gain setting, is approximately 20GHz in simulation (RCC extracted). In each sub-ADC, a VGA follows the T&H on a 1.2V supply to accept the high common-mode from the sampler. All other circuits operate from a 0.9V supply. After the VGA is an MSB comparator, a differential folding stage, and a 5b flash. The 1b folding stage rectifies the signal based on the MSB decision, reducing the number of comparators by half, hence saving power. Each comparator in the 5b flash can be individually disabled (clock-gated) for non-uniform quantization. Note that the 1b folding stage does not affect the non-uniform quantizing effect of the 5b flash since the sampled input distribution is almost symmetric in a differential link. This fact also reduces the search space for the optimal ADC threshold levels. The flash utilizes a Wallace Tree encoder and a resistor ladder for reference generation. In order to interpret the output of

the flash when non-uniform quantization is employed, an encoder is needed and can be constructed as a LUT [4] (simulated power of  $\sim 1.5mW$  per subADC in this technology). In this prototype, the encoder and subsequent digital equalizer is implemented off-chip in software. In order to power scale the ADC for channels not requiring the full 6b resolution, a fast greedy-search algorithm is used. Starting with a uniformly quantized 6b ADC, the back-end digital equalizer (EQ) taps are adapted using an LMS algorithm. Next each quantization level is removed in turn by deactivating one comparator at a time and the EQ taps are re-converged. The level whose removal causes the smallest increase in BER is removed. This procedure is repeated until a power consumption or BER target is reached.

Prototypes are fabricated in TSMC 16nm FinFET CMOS. The TX occupies an active area of  $500\mu m \times 180\mu m$ , while the RX occupies an active area of  $650\mu m \times 250\mu m$  (Fig. 6.5.7). TX measurements are shown in Fig. 6.5.3. The PAM-4 PRBS15 eye, with package and PCB losses de-embedded at 64.375Gb/s, shows that the TX achieves an RLM  $> 0.99$  at  $1V_{ppd}$  full swing. Linearity control is possible to increase the lower eye heights by  $\sim 20mV$  each. Clock pattern jitter decomposition at 32.1875Gb/s shows an RJ of  $162fs_{rms}$  and a TJ of 2.82ps at  $1e-12$ . The TX draws 89.7mW including clock distribution (simulated 26% of total) from a 1.2V supply.

The RX AFE measurement results are shown in Fig. 6.5.4. An SNDR of 27.8dB at a Nyquist input frequency of 16GHz is observed after static nonlinearity compensation is performed using an off-chip DAC to tune comparator thresholds. Peak-to-peak gain and offset mismatch between the 8 sub-ADCs up to 2.2% and 2.18LSB respectively are corrected with a combination of coarse analog correction (on-chip), and fine digital correction (off-chip) based on a single-tone low-frequency calibration. DCC suppresses the skew tone to  $-45dBFS$ . The CTLE sampler measurements exhibit  $\sim 6dB$  boost with the 2<sup>nd</sup> pole at  $> 20GHz$ . The complete RX-AFE consumes 283.89mW (including clocking, excluding retimer) from 0.9V/1.2V supplies. Link measurements are shown in Fig. 6.5.5 using 3 channels with different loss at 16GHz: A) 8.6dB; B) 13.6dB; and C) 29.5dB. In performing these measurements, the phase is frozen and swept using a phase interpolator on the TX side. For A, the TX equalization ( $-13.6\%$  precursor and  $-9.1\%$  postcursor) in combination with RX CTLE is enough to open the eye as shown. Using the ADC in 2b mode (as a slicer) is sufficient to obtain a BER  $< 1e-6$ , yielding a power efficiency of 3.3pJ/b (TX+RX, excluding CDR, PLL). For C, the eye is completely closed at the input to the ADC. Hence, the entire 6b uniform ADC is used in conjunction with CTLE at maximum boost and TX equalization ( $-13.6\%$  precursor and  $-22.7\%$  postcursor) to achieve a BER  $< 1e-4$  with a 16-tap FFE (off-chip), yielding 6.2pJ/b (excluding CDR, PLL, and equalizer). Channel B is used to investigate non-uniform quantization targeting a BER of  $1e-5$  (limited by ADC memory acquisition time). The ADC is first calibrated as a uniform 6b quantizer with a 6-tap FFE. Comparators are then disabled using a greedy search with BER as the optimization metric. The final 31 non-uniform levels yield a BER of  $4.2e-6$  at 4.26pJ/b compared to a BER of  $2.8e-5$  using the ADC in 5b uniform mode. Comparisons with recent state-of-the-art TX and RX are shown in Fig. 6.5.6. This TX achieves the best jitter, RLM and power efficiency compared to [1,2,5]. The ADC-based RX-AFE is comparable in power efficiency to existing work and its folding flash architecture allows for aggressive power scaling and non-uniform quantization that the SAR architectures used in [1,2] do not.

### Acknowledgments:

The authors would like to acknowledge the financial and logistic support provided by Huawei, especially Dustin Dunwell, David Cassan, and Davide Tonietto, and the Huawei layout team, especially Muhammad Ali Khan, Diana Ilieva, Mark Roberts, and Trevor Monson.

### References:

- [1] Y. Frans, et al., "A 56-Gb/s PAM4 Wireline Transceiver Using a 32-Way Time-Interleaved SAR ADC in 16-nm FinFET," *IEEE JSSC*, vol. 52, no. 4, pp. 1101-1110, Apr. 2017.
- [2] K. Gopalakrishnan, et al., "A 40/50/100 Gb/s 4-PAM Ethernet Transceiver in 28nm CMOS," *ISSCC*, pp. 62-63, Feb. 2016.
- [3] E. H. Chen, et al., "Power optimized ADC-based serial link receiver," *IEEE JSSC*, vol. 47, no. 4, pp. 938-951, Apr. 2012.
- [4] Y. Lin, et al., "A Study of BER-Optimal ADC-Based Receiver for Serial Links," *IEEE TCAS-I*, vol. 63, no. 5, pp. 693-704, May 2016.
- [5] G. Steffan, et al., "A 64Gb/s 4-PAM Transmitter with 4-Tap FFE and 2.26pJ/b Energy Efficiency in 28nm CMOS FDSOI," *ISSCC*, pp. 116-117, Feb. 2017.
- [6] Y. Duan, E. Alon, "A 12.8 GS/s Time-Interleaved ADC With 25 GHz Effective Resolution Bandwidth and 4.6 ENOB," *IEEE JSSC*, vol. 49, no. 8, pp. 1725 - 1738, Aug. 2014.

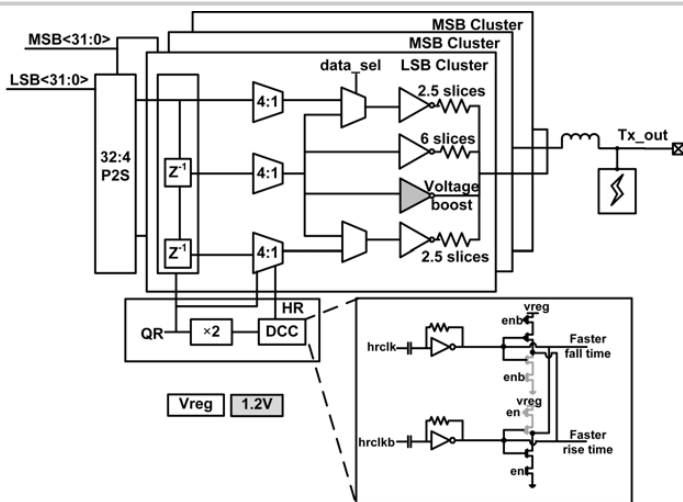


Figure 6.5.1: TX block diagram: 3 identical driver clusters, of which 2 driven by the MSB serializer. A DCC circuit provides adjustment (250fs res.) by controlling delay on the rising edge of clock and falling edge of clock.

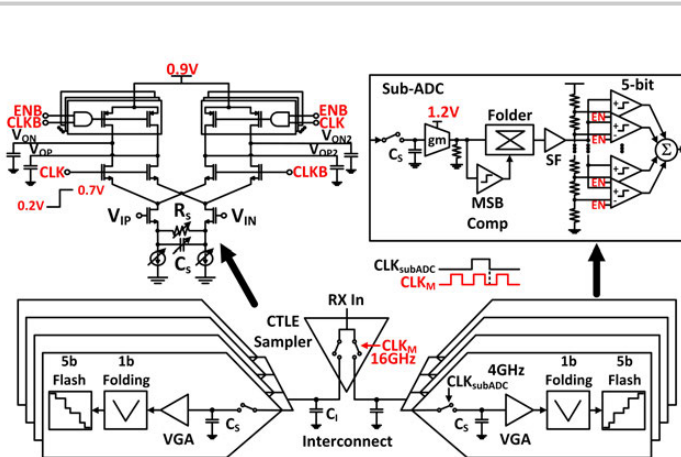


Figure 6.5.2: RX block diagram: 32GS/s 8-way time-interleaved ADC with CTLE half-rate sampler, sub-ADC with 1b folding followed by 5b full flash with individual enables for each comparator.

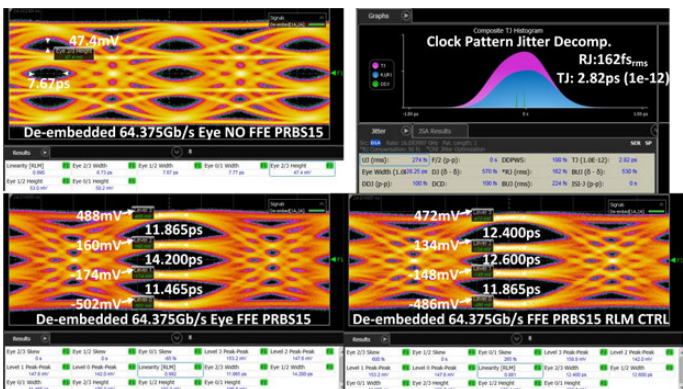


Figure 6.5.3: TX measurements: De-embedded PAM-4 PRBS15 eyes without and with FFE weights [-1.5, 30, -1.5], and with linearity control; jitter decomposition with clock pattern.

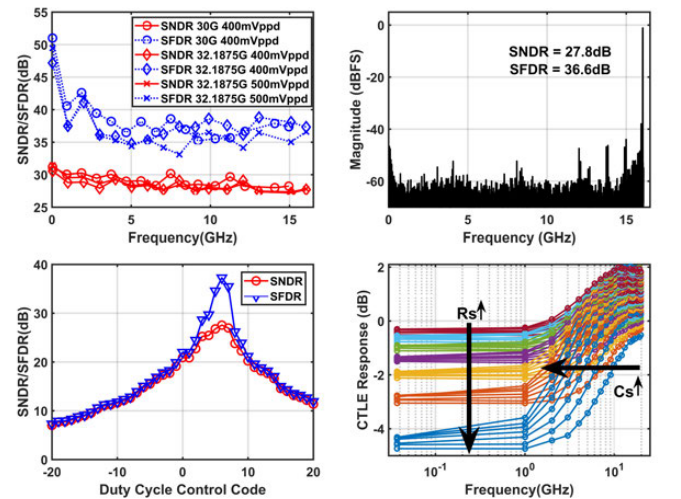


Figure 6.5.4: RX measurements: SNDR/SFDR; 16k point FFT with Nyquist frequency input; front-end sampling CLK DCC CTRL; CTLE responses relative to flat setting.

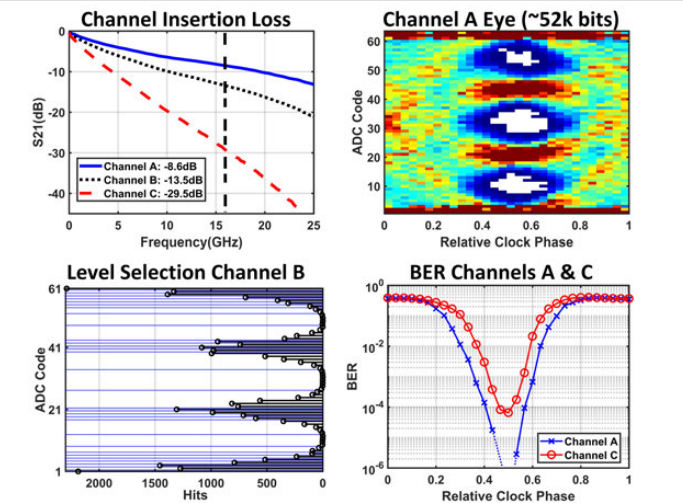


Figure 6.5.5: Link measurements: Insertion loss; Channel A eye (52k bits) at ADC output; Channel B ADC output histogram (black) and optimized 5b non-uniform thresholds (blue); BER bathtub for channels A and C.

	[1]	[2]	[5]	This work
Data Rate (Gb/s)	56	64	64	64.375
RX ADC Res (bits)	8	8	-	6
RX AFE ENOB at Nyquist (bits)	4.9	4.9**	-	4.31
RX AFE Power (mW)	370*	-	-	100@ChA 184.9@ChB 283.9@ChC
TX Swing (V <sub>pp-diff</sub> )	1.2	1.4	1.2	1
TX RJ (fs <sub>rms</sub> )	200	240	290	162
TX RLM (%)	97	-	94	99
TX Power (mW)	140	-	145	89.7
Supplies (V)	0.9/1.2/1.8	0.9/1.2	1.0/1.2	0.9/1.2
Chip Area (mm <sup>2</sup> )	2.8	30.87***	-	TX: 0.09 RX: 0.1625
Technology	16nm FinFET	28nm	28nm FDSOI	16nm FinFET

\*Including retimer, \*\*at 10GHz in figure, \*\*\*total area including 2 RX/TX/PLL & DSP.

Figure 6.5.6: Comparison Table: this transceiver is the only one which allows for non-uniform quantization and has the best power scaling capability.

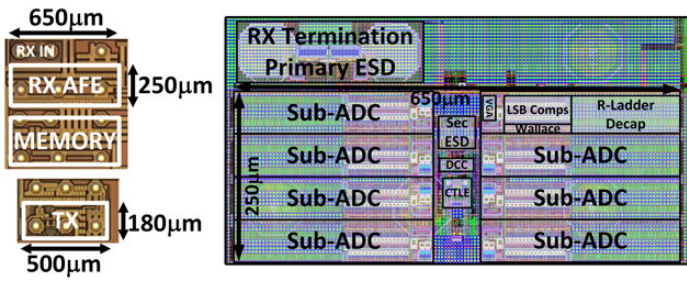


Figure 6.5.7: Die Photos: TX active area 500µm×180µm, RX AFE active area 650µm×250µm; RX-AFE layout view with sub-blocks annotated.