

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

Luke Wang^{1b}, *Student Member, IEEE*, Yingying Fu, Marc-Andre LaCroix, *Member, IEEE*,
Euhan Chong, *Member, IEEE*, and Anthony Chan Carusone^{1b}, *Senior Member, IEEE*

Abstract—A 64-Gb/s 4-pulse-amplitude modulation (PAM) transceiver fabricated with a 16-nm fin field effect transistor (FinFET) technology is presented with a power consumption that scales with link loss. The transmitter (TX) includes a three-tap feed-forward equalizer (FFE) (one pre and one post) achieving a level separation mismatch ratio (RLM) of 99% and a random jitter (RJ) of 162-fs rms. The maximum swing is 1.1 V_{ppd} at a power consumption of 89.7 mW including clock distribution from a 1.2-V supply, corresponding to 1.39 pJ/bit. The receiver analog front end (RX-AFE) consists of a half-rate (HR) sampling continuous-time linear equalizer (CTLE) and 6-bit flash (1-bit folding) analog-to-digital converter (ADC) capable of non-uniform quantization. The non-uniform thresholds are selected based on a greedy search approach which allows the RX to reduce power at low channel loss in a highly granular manner and achieves better bit error rate (BER) than a uniform quantizer. For a channel with -8.6 -dB loss at Nyquist, ADC can be configured in 2-bit mode, achieving $BER < 1e-6$ at an RX-AFE power consumption of 100 mW. For a -29.5 -dB loss channel, the RX-AFE consumes 283.9 mW and achieves a $BER < 1e-4$ in conjunction with a software digital equalizer. For a -13.5 -dB loss channel, a greedy search is used to optimize the quantization threshold levels, achieving an order of magnitude improvement in BER compared to uniform quantization.

Index Terms—4-pulse-amplitude modulation (PAM), adaptive algorithms, analog-to-digital conversion, equalizers, fin field effect transistors (FinFETs), greedy algorithms, receivers (RXs), transceivers, transmitters (TXs).

I. INTRODUCTION

AS DATA rates continue to increase, frequency-dependent channel loss is making it difficult for wireline links to achieve acceptable bit error rates (BERs). To decrease the insertion loss (IL) at the Nyquist frequency, many wireline links are transitioning from nonreturn to zero (NRZ) [2-pulse-amplitude modulation (PAM)] to 4-PAM modulation including standardized links such as IEEE 802.3 bs [1] and OIF CEI-56G-PAM4 [2]. Traditional mixed-signal receivers (RXs) are no longer capable of handling channels with

>20 -dB loss at the Nyquist frequency and >2 mV_{rms} of crosstalk with 4-PAM [3], which require complex equalization. Thus, an analog-to-digital converter (ADC)-based RXs, where a high-speed ADC digitizes the received waveform and digital adaptive equalization is applied, have become increasingly popular [4]–[6]. Fig. 1 shows the structure of a typical mixed-signal RX and an ADC-based RX. The analog front end may look deceptively similar; however, the high-frequency boost requirement and therefore power consumption of the continuous-time linear equalizer (CTLE) may be relaxed in the ADC-based RX due to the addition of extensive digital equalization after the ADC. In addition to the CTLE, the mixed-signal RX usually incorporates a decision-feedback equalizer (DFE). However, it is difficult to implement multi-tap DFEs at a data rate ≥ 56 Gb/s [3]. In the ADC-based RX, the digital signal processor (DSP) consists of many taps (>10) of feed-forward equalizer (FFE) and a few taps (1–2) of DFE [4], [5]. The FFE taps are simpler to implement and consume less power than the DFE taps; however, they require the ADC resolution to be high in order to be effective. For 4-PAM RXs covering up to 30 dB of loss, resolutions of 7–8 bits are generally used [4]–[6], necessitating a successive approximation register (SAR) architecture for good power efficiency. In the context of link power scaling, the power consumption of an ADC with a SAR architecture scales roughly linearly with resolution, and therefore, insignificant power is saved at lower channel loss. For instance, in [5], only 10% of the RX analog front-end (RX-AFE) power is saved when switching from a 7-bit to a 3-bit resolution. Since the majority of data-center links between servers and switches are low loss (<20 dB), an RX-AFE that has good power scaling capability in this loss range may be preferable. In addition, at this loss region, the ADC resolution required is 6-bits or less [7], thus opening new possibilities for the ADC architecture design.

II. POWER SCALABLE ADC ARCHITECTURE FOR VSR-MR

For a typical 4-PAM mixed-signal RX with D -tap DFE, if loop unrolling is used, $3(4^D)$ comparators are necessary. Thus, as shown in Fig. 2(a), even a one-tap loop-unrolled DFE requires 12 comparators. The structure resembles that of a flash ADC with non-uniform threshold levels clustered around the lower (DL), middle (DZ), and higher (DH) slicing levels. Since the flash architecture's power consumption scales exponentially with its resolution and its power efficiency at

Manuscript received June 6, 2018; revised September 9, 2018; accepted October 14, 2018. Date of publication November 8, 2018; date of current version January 25, 2019. This paper was approved by Associate Editor Jack Kenney. This work was supported by Huawei Canada. (*Corresponding author: Luke Wang.*)

L. Wang and A. Chan Carusone are with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada (e-mail: luke.wang@isl.utoronto.ca).

Y. Fu is with Huawei Canada, Markham, ON L3R 5A4, Canada.

M.-A. LaCroix and E. Chong are with Huawei Canada, Kanata, ON K2K 3J1, Canada.

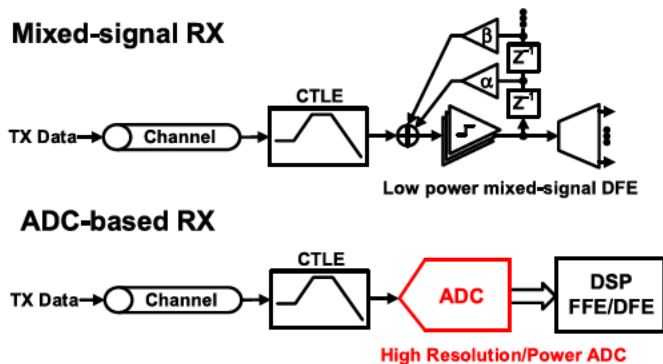


Fig. 1. Typical mixed-signal RX architecture used predominately for low-loss (<20 dB) channels in chip-to-chip and chip-to-module applications, and ADC-based RX architecture used in high loss (20–30 dB) chip-to-backplane/cable applications.

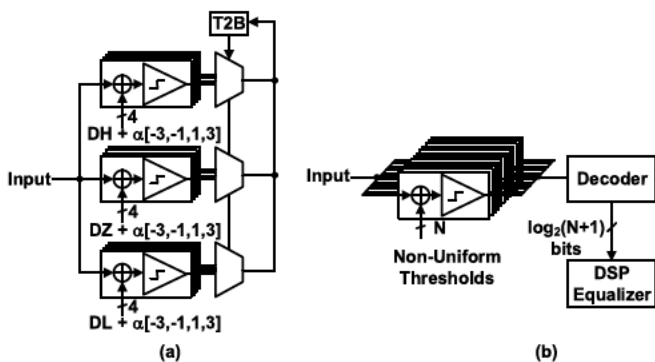


Fig. 2. (a) One-tap 4-PAM loop-unrolled DFE. (b) Generic non-uniform quantizing flash ADC.

6-bit resolution and below (as required by links at 20-dB loss or less) is comparable to a SAR, it is a good choice for a power scalable RX solution. Thus, a flash-based architecture with variable non-uniform thresholds in conjunction with a reconfigurable equalizer is preferred as shown in Fig. 2(b).

A. Background and Prior Art

RXs with non-uniform quantization and DSP realizing approximate DFEs have been applied to NRZ in [8]–[10]. In [8] and [9], the result is called a reduced-slicer partial-response DFE (RS-PRDFE) in which M comparators are used, where $M < 2^D$, to approximate a D -tap DFE. As the name implies this is effectively a loop-rolled DFE where similar intersymbol interference (ISI) values are grouped together such that some comparators/thresholds cover more than one ISI level. In [10], a look-ahead DFE is used to choose a decision sequence, termed sequence DFE; however, this is done in a redundant way that the choice does not depend on the result of a single comparison. Both of these works apply only to a DFE equalizer. In the 4-PAM case, [7] uses a variable threshold flash ADC to change the threshold levels dynamically during conversion. An extra set of comparators sample the data at its edge location and use this information to dynamically adjust the thresholds (R-ladder references) for the center sample. This method is, however, constrained by the

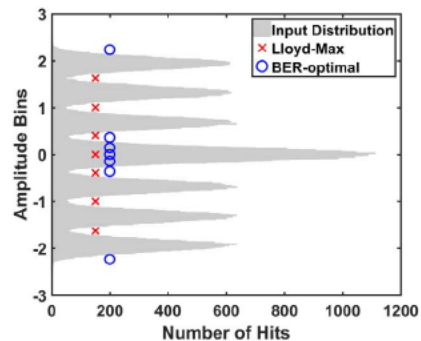


Fig. 3. NRZ input distribution for simulated channel [0.31, 1, 0.63] and a SNR = 20 dB, with a one pre-two post-tap FFE minimum mean squared error (MMSE) equalizer. LM levels (red crosses) are shown in contrast to BER-optimal levels (blue circles).

reference settling time requirement and therefore also power consumption in reference switching. In [11], a least mean squares (LMS) gradient descent is used to tune the thresholds of a flash converter. This, however, requires a high-resolution Digital-to-Analog Converter (DAC) in the comparator of the converter: in [11], an 8-bit DAC is used for a 4-bit ADC. It also has the potential problem of divergence.

Non-uniform quantization has been extensively studied more generally. For example, the Lloyd–Max (LM) algorithm [12] can be used to select a set of threshold levels that minimize quantization noise power for a signal with known statistics. Intuitively, this means concentrating the levels where the input probability density function (PDF) has the highest value. However, it can be shown that in the presence of ISI and an equalizer, the LM quantizer does not allow a wireline link to achieve the lowest possible BER [13]. An example is shown in Fig. 3, where binary (± 1) symbols are applied to a baud-rate discrete-time channel $0.31 + z^{-1} + 0.63z^{-2}$ with additive white Gaussian noise (AWGN) at an SNR of 20 dB resulting in the baud-rate sampled PDF plotted in gray. A set of seven threshold levels are chosen for the quantizer, which is followed by a three tap (one precursor, two postcursor) FFE. The BER-optimal levels are found by exhaustively searching all combinations of 63 threshold levels uniformly spaced over the entire input range. Note that due to the symmetry of the PDF, the search can be simplified to consider only threshold levels that are symmetrically arranged around zero, a fact used later in our system architecture. In this example, a large difference is evident between the LM levels and BER-optimal levels. In the BER-optimal case, the levels are more concentrated around zero, where due to ISI and noise, the 0 and 1 bits are ambiguous. At the lower and upper amplitude extremes, a hard decision is possible, and therefore, more quantization error can be tolerated.

Exhaustive search is, however, impractical in a real implementation since the number of combinations of levels that needs to be tested is prohibitive. Assuming a symmetric input distribution, seeking the K best threshold levels from among all the levels in an N -bit ideal ADC, $\binom{2^{N-1} - 1}{K}$ combinations must be tested. For instance, for $N = 6$ and $K = 7$, 2 629 575 combinations need to be tried. For each trial,

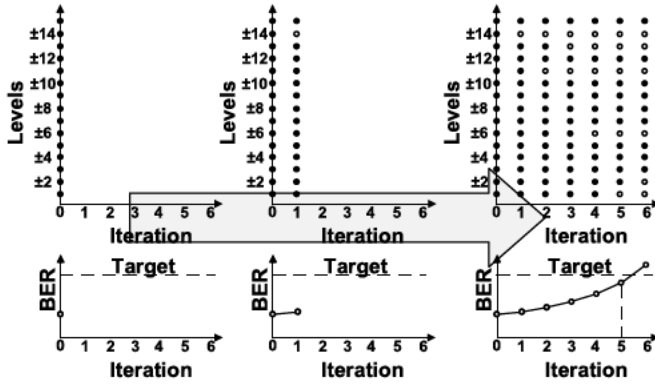


Fig. 4. Greedy search progression illustration: starting from a symmetric 5-bit ADC, threshold levels are removed until the BER limit is no longer satisfied.

the BER is measured, implying a prohibitively long calibration cycle.

B. Greedy Search Approach

To shorten the search time, a greedy search is employed in this paper. The iterative search works by selecting a subset of the candidate levels at each step and using that subset as candidates for the subsequent iteration. Fig. 4 shows the progression of a greedy search beginning with 31 threshold levels (5-bit ADC resolution) symmetrically arranged so that levels may be removed in pairs: $\pm 15, \pm 14, \dots$. The search begins with all threshold levels active and the link operating at a BER well below target. In the first iteration, after removing each pair of levels one at a time (15 trials), it is observed that removal of levels ± 14 causes the smallest increase in BER. Hence, levels ± 14 are deactivated and a tolerable increase in BER results. In the next iteration, 14 trials are made removing an additional pair of levels and identifying ± 11 for deactivation. This process is repeated until iteration 6, where after removing levels ± 2 , the BER exceeds the target. Therefore, the search is terminated, and the set of levels corresponding to iteration 5 may be used.

The number of required trials using greedy search is far less than in an exhaustive search but does not necessarily result in the global optimum. In order to see the performance degradation compared to an exhaustive search, a 4-PAM RX using a non-uniform quantizer was simulated. Using a baud-rate discrete-time channel $0.12 + z^{-1} + 0.49z^{-2}$, with AWGN at 30-dB SNR and one precursor plus two postcursor-tap FFE, the quantizer is reduced from 5-bit (31) uniformly spaced threshold levels to 15 levels. Fig. 5 shows the number of combinations satisfying each BER limit, where the total number of combinations, assuming symmetry, is $\binom{15}{7} = 6435$. Among these, one combination achieves a BER of 10^{-5} , while the greedy search achieves a BER of 20×10^{-5} , which places it among the best 20 combinations, well above the 99.5th percentile. Moreover, this BER is an order of magnitude lower than for a uniform 4-bit quantizer (also having 15 threshold levels), which achieves a BER of 250×10^{-5} . Note that the simulation parameters are chosen for a relatively high BER

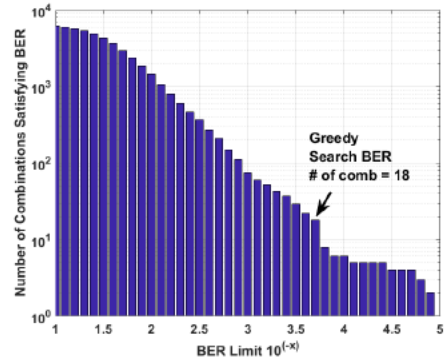


Fig. 5. Threshold level selection simulation for 4-PAM input for channel $[0.12, 1, 0.49]$ and a SNR = 30 dB, with a one-pre/two-post tap FFE MMSE equalizer.

TABLE I
SIMULATION OF GREEDY SEARCH PERFORMANCE FOR REPRESENTATIVE CHANNELS WITH 4-PAM (SNR = 30 dB)

N_{FFE} (N_{PRE})	N_{DFE}	Loss@ Nyq (dB)	BER 5-bit Uniform	BER 4-bit Uniform	BER 4-bit Greedy	BER 4-bit LM
9(2)	1	25	5.5e-6	2.1e-2	7.4e-3	1.1e-2
4(1)	0	20	3.5e-6	1e-3	4.87e-4	7.8e-4
3(1)	0	15	1.1e-6	3.9e-5	1.2e-5	2.3e-4
1(0)	1	10	2e-7	1.62e-5	2e-7	1.56e-5

so that an exhaustive search of all combinations is possible. In practical scenarios, with more combinations and lower BER, simulation of an exhaustive search is impractical.

In order to examine the performance of greedy search for 4-PAM input in more detail, several representative channels from measured data made available by the IEEE 802.3cd task-force [14] were used. The SNR was set to 30 dB, the quantizer was scaled from 5 to 4 bit using both greedy search and LM, and an appropriate equalizer with N_{FFE} FFE taps (N_{PRE} pre-cursor taps) and N_{DFE} DFE taps was used to achieve a BER practical in the simulation. For a fair comparison, the LM levels are quantized to the same precision as the greedy search levels. The results are summarized in Table I. Greedy search always achieves better BER compared to the uniform and LM quantizer. For higher channel loss, the advantage of greedy search compared to LM quantizer begins to fade since a long FFE is used for a Gaussian-like input distribution. Note that the channel loss in Table I corresponds to a system without any equalization before quantization, therefore leading to a pessimistic photograph for the applicability of greedy search as a function of channel loss. In the low-loss cases, the greedy search levels can indeed be quite different from the LM levels as shown in Fig. 6. The use of DFE, here, leads to this optimal level distribution since the thresholds should be concentrated around the optimal slicing levels (eye centers) \pm the predicted residual ISI values as in the case of a loop-unrolled DFE. This is consistent with the analysis presented in [8]. In the LM case, the levels are of course concentrated away from the eye centers and instead near where the most samples are.

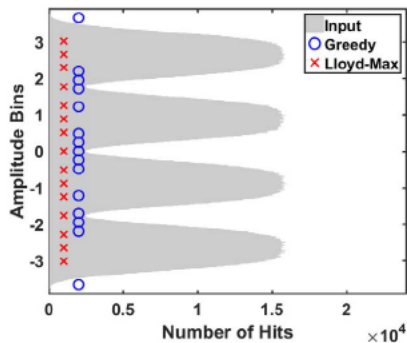


Fig. 6. Greedy search levels for simulated 10-dB loss case with one tap DFE equalizer showing level placement in contrast to minimum quantization error (LM) case.

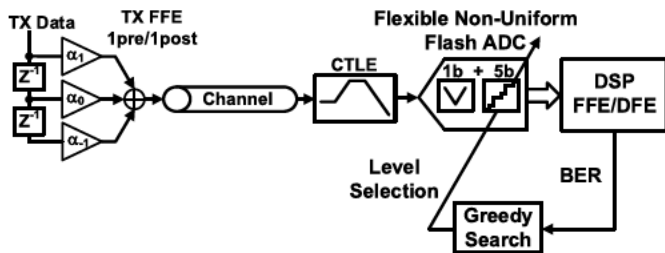


Fig. 7. System level block diagram of 64-Gb/s transceiver.

C. Proposed System Architecture

Fig. 7 shows a system level diagram of the proposed 64-Gb/s transceiver. The transmitter (TX) is a voltage mode TX with one precursor and one postcursor-tap FFE. The RX has a CTLE incorporating a half-rate (HR) sampler as will be described in Section III. The CTLE is then followed by a 32-GS/s 6-bit ADC implemented as a 1-bit folding stage followed by a 5-bit full flash. Vertical symmetry of the input PDF is assumed for the differential input; hence, the non-uniform threshold selection may be applied after the 1-bit folding stage to both reduce power consumption of the flash and reduce the search space of the threshold levels. 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 look-up table (LUT) with low area/latency overhead [11] consuming a simulated power of 1.5 mW per sub-ADC in this technology. In this prototype, the DSP including the encoder and feedback loop using greedy search is implemented off-chip in software. The DSP computes a BER estimate based on the Pseudo-random binary sequence (PRBS) checker output after the data are equalized. In a real 4-PAM system, a forward error correction (FEC) unit may provide this BER estimate. The BER is then used to guide the greedy search to de-activate the correct level for each iteration. In this paper, the equalizer coefficients are re-adapted for each trial in the greedy search in order to capture any co-dependence between the quantizer threshold level selection and the equalizer coefficients. This paper does not address timing recovery; however, since the ADC is power scaled from full uniform resolution, the initial timing recovery can be the same as traditional ADC-based RXs. Once locked and a greedy search begins, it will consider

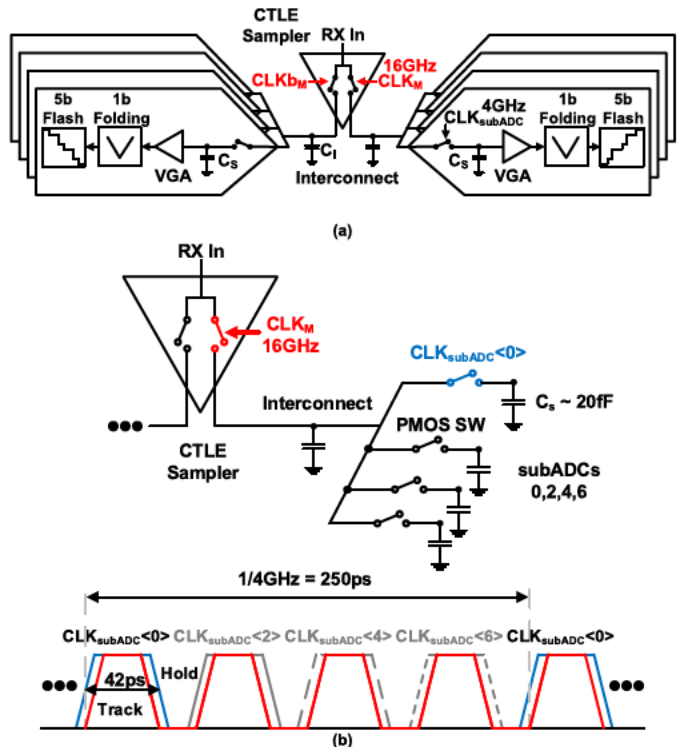


Fig. 8. (a) RX-AFE architecture: CTLE sampler driving eight-way TI 6-bit ADC. (b) Sampling structure with front-end sampler passing samples to sub-ADC.

the impact of timing recovery upon BER in its optimization to maintain a targeted BER.

III. RECEIVER DESIGN

The RX-AFE consists of a front-end sampling CTLE and an 8-way time-interleaved 32-GS/s flash ADC. As shown in Fig. 8(a), the CTLE sampler drives two banks of four sub-ADCs, one on each side. It functions as a two-way time-interleaved sampler with peaking, requiring differential 16-GHz CMOS level clocks. Clock skew is typically a major issue in the design of time-interleaved ADCs [15]. The two-way sampler simplifies this to duty-cycle correction (DCC) of the 16-GHz clock. Fig. 8(b) shows how the sample is passed from the front-end sampler to the sub-ADCs (only one side is shown for simplicity). During track mode, both the master switch inside the front-end sampler (red) and the sampling switch inside the sub-ADC (blue) are ON. As long as CLK_M falls before CLK_{subADC} , the skew will only be the duty-cycle distortion (DCD) of CLK_M . The width of CLK_{subADC} needs to be wider than CLK_M (i.e., >31.25 ps) across process-voltage-temperature (PVT) variations, but not so wide that it overlaps with the pulses from other sub-ADCs (i.e., <62.5 ps), which would lead to crosstalk between sub-ADCs and degraded bandwidth (BW). In this paper, a tracking width of 42 ps is used for the sub-ADCs, and CLK_M and CLK_{subADC} are aligned by adjusting a delay line in the master clock path. There is no need for a separate delay adjustment on each sub-ADC clock as long as the skew $\lesssim 11$ ps. A sampling capacitor of approximately 20 fF comprised of layout

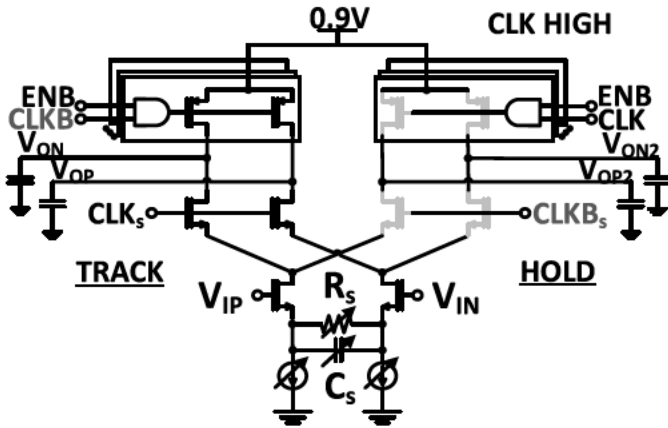


Fig. 9. Two-way time-interleaved CTLE sampler when CLK is high: left side is in track mode and right side is in hold mode.

parasitics and the input capacitance of the next stage is used. The BW of the sampling network was simulated after extraction, including interconnect parasitics, to be >20 GHz.

A. Front-End Sampler Design

Fig. 9 shows the schematic of the front-end sampler used in this paper. The sampler works as a two-way time-interleaved front end by allowing one side to track the input, while the other side is holding the input. When tracking, the transistors on the bottom form a cascode, which provides a high impedance. Thus, the PMOS triode load, which has the programmable width and a maximum resistance of 50Ω , together with the capacitance at the output determines the BW. This sampler was first proposed in [16] and was shown to be better than having the series connection of a buffer and a switch when the target BW approaches a fraction of the transit frequency f_t . In this paper, the differential pair trans-conductor is resistively and capacitively degenerated to provide a 6-dB boost at 16 GHz. The clocking structure is also modified as will be discussed in the following.

One challenge of the sampler is the clock generation for the PMOS load and NMOS cascode. For the NMOS cascode transistor, the clock swing must be reduced to keep it in saturation, while the PMOS uses a full-swing clock to reduce its size for the same ON-resistance. Any timing mismatch between the clock driving the cascode and the one driving the load will cause common mode (CM) problems. For instance, as shown in Fig. 10(a), if the cascode device turns off first, the output CM (V_{ocm}) will be pulled high. To remedy these issues, Fig. 9(b) shows the added alignment control between clocks (± 4 ps in 1-ps resolution) to stabilize the CM and the CMOS-style reduced swing buffer which provides a similar rise/fall time to the full swing buffered clock. By increasing the size of M_n , the pull-down is reduced and the low-level increases, while increasing the size of M_p , the pull-up is reduced and the high-level decreases. In order to control the alignment during testing, the average drain voltages V_{ds} of the bias devices of the sampler are monitored through a test-bus connection. The control code is then tuned to maximize signal to noise and distortion ratio (SNDR) while ensuring

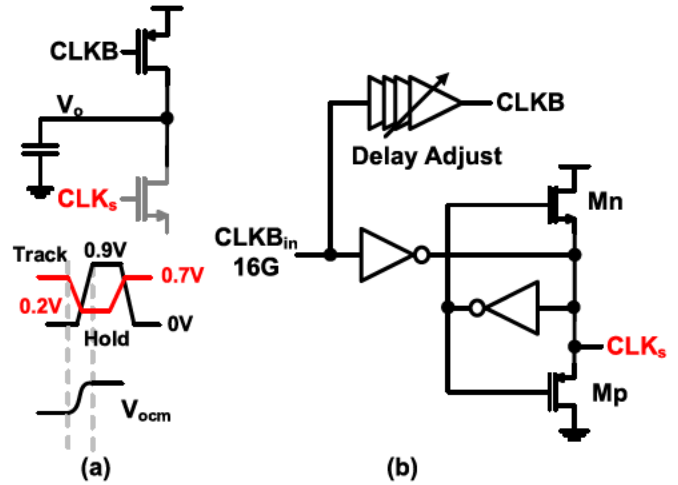


Fig. 10. Clock generation for CTLE cascode sampler. (a) CM shift problem due to clock timing mismatch. (b) Clock phase alignment to prevent CM shift, and reduced swing buffer for generating clock for cascode NMOS.

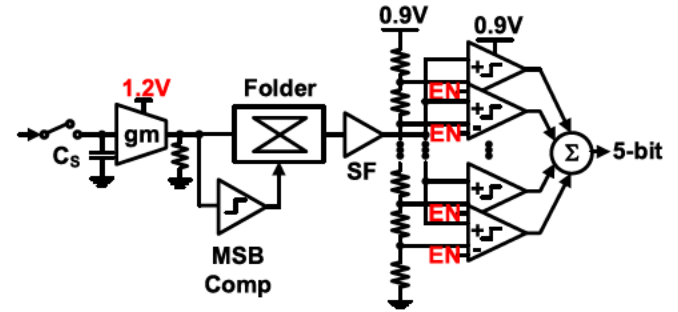


Fig. 11. Sub-ADC circuit architecture: gm-stage implementing VGA followed by 1-bit folding +5-bit full flash.

both V_{ds} and sub-ADC variable gain amplifier (VGA) range stays within design specifications.

B. Sub-ADC Design

As shown in Fig. 11, each 4-GS/s 6-bit sub-ADC, a modification of [17], consists of a PMOS sampling switch, a trans-conductor (gm) stage which also acts as a VGA, a 1-bit folding stage which is triggered by the MSB comparator decision, and a 5-bit full flash ADC buffered by a PMOS source follower (SF) followed by a Wallace tree adder. The VGA is the only circuit in the RX-AFE which uses a 1.2-V supply in order to accept the high output CM (~ 0.7 V) of the CTLE sampler. A 0.9-V supply is used for all other blocks. Each comparator in the full flash consists of a clock-gated enable (EN) so that a subset may be de-activated for non-uniform quantization to save power.

Fig. 12 shows the complete timing diagram. A divide-by-4 generates eight phases of a 50% duty-cycle 4-GHz clock from the 16-GHz master clock. All clocks within each sub-ADC are generated from two adjacent phases of the 4-GHz clock which are routed to each sub-ADC. These two adjacent phases in conjunction with a variable delay (not present in [17]) allow the track pulsewidth to be set $>1/32$ GHz. All other clocks are generated with simple

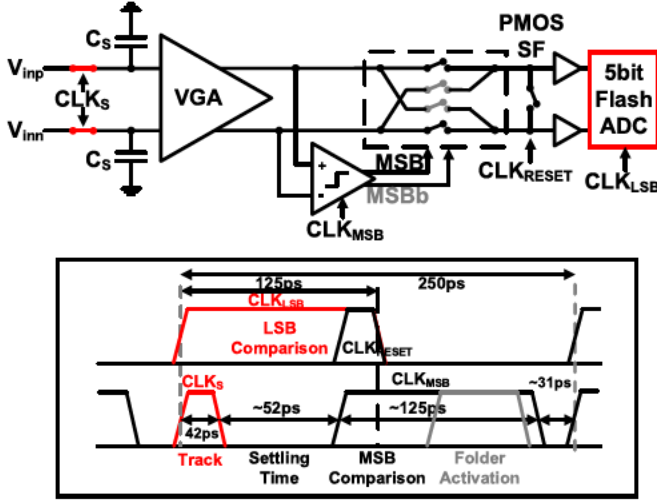


Fig. 12. Sub-ADC timing diagram: folding switches are used to pipeline the conversion.

combinatorial logic as well. The folding switches are activated based on the MSB comparator decision and simply exchange the differential signal lines if the differential input is negative. During the MSB comparator reset, all switches are open. Thus, these switches are also used to pipeline the conversion as highlighted in red; the tracking window, the settling time of the VGA, as well as the initial triggering of the MSB comparator all take place at the same time as the LSB comparison of the previous sample. This works as long as the folding switches are inactive during the LSB comparison, which is guaranteed by the design. This pipelining is different from prior art [18] in which an additional sample and hold (S&H) is added between the input to the MSB comparator and the VGA. In order to reduce ISI, the input to the SF is reset at the end of the LSB comparison cycle. This improves the SNDR significantly, especially when the input is at $f_{\text{subADC}}/4$. This degradation is observable in [17] without the reset. Note that this is not the Nyquist frequency $f_{\text{subADC}}/2$ because after folding the highest frequency content is doubled. The reset does, however, increase the meta-stability probability of the LSB comparator as it overlaps the decision time as shown in Fig. 12.

A double-tail latch [19] was used for both the MSB and LSB comparators. In this technology, the comparators achieve a simulated regeneration time constant of approximately 3 ps after layout extraction. The LSB comparators (with NMOS input) are designed for an input-referred thermal noise of 1.2 mV_{rms} which is less than quantization noise at a sub-ADC full scale of 400 mV_{ppd}. For offset calibration purposes, a 5 bit (2-bit thermometer and 3-bit binary) MOS-capacitor capacitive DAC (CDAC) is included in each comparator. It has minimal impact on the input-referred noise of the comparator at the cost of added propagation delay. Due to the high thermal noise in fin field effect transistor (FinFET) technologies (~2× compared to planar), the comparators cause significant kickback onto the SF due to the larger device size. In order to alleviate this, two techniques are used as shown in Fig. 13. In Fig. 13(a), the pre-amplifier is augmented so that M1 and M2 reset the source node so that the sampled voltage does

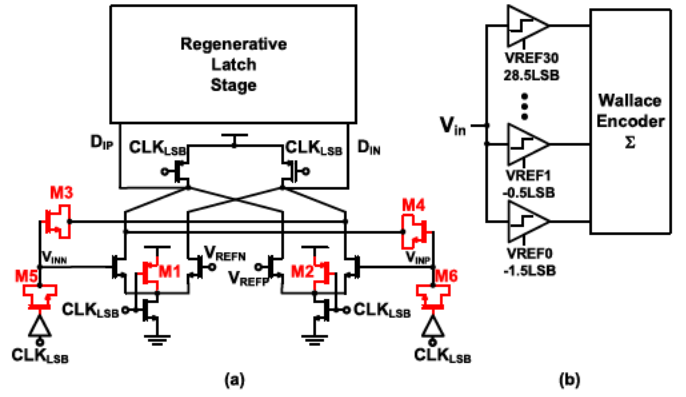


Fig. 13. Kickback reduction for comparators. (a) Augmenting the dynamic pre-amplifier. (b) Shifting the reference ladder references for systematic offset.

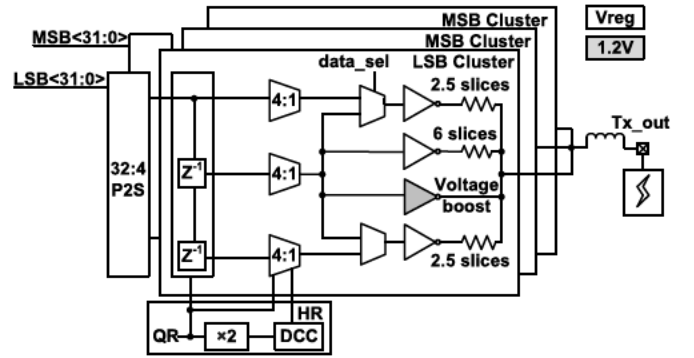


Fig. 14. TX block driver clusters, two driven by MSB, one driven by LSB, implementing a slice-based one pre-/one post-tap FFE.

not drift, M3 and M4 convert differential kickback to CM, and M5 and M6 are added at the gate of the input devices to supply charge [20]. Since the input to the full flash is folded, only positive differential voltages cause unidirectional kickback. Thus, as shown in Fig. 13(b), the references to the comparators are shifted by two LSBs to account for this systematic shift.

IV. TX DESIGN

The TX shown in Fig. 14 consists of three identical driver clusters driven by separate 32–4 serializers. The LSB serializer drives one driver cluster while the MSB serializer drives two clusters. In NRZ mode, identical data streams are applied to both MSB and LSB serializers. Each driver cluster consists of 11 source-series terminated (SST) slices, six of which are 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 4-PAM mode. In parallel with the 11 SST voltage-mode slices, a voltage boost stage is connected as shown in Fig. 14. This voltage booster is a current mode stage similar to [21]. The additional current it injects to the output nodes is capable of increasing the output voltage swing by 100 mV_{ppd} (10%) to a maximum of 1.1 V_{ppd}. The slice-based design enables termination calibration and facilitates eye pre-distortion where relative weights of MSB and LSB slices can be tuned to

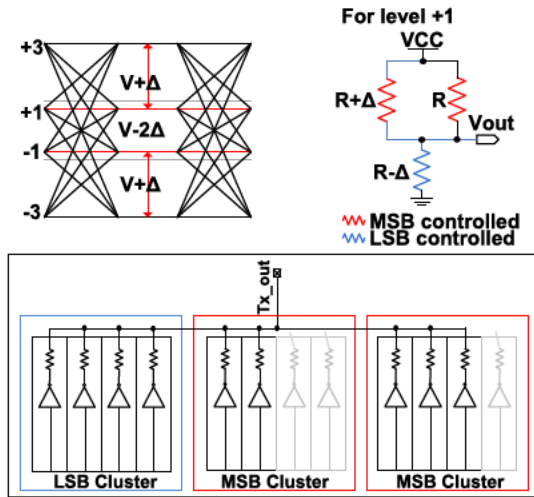


Fig. 15. TX predistortion adjustment: the center eye vertical opening is decreased, example generation shown for +1 level.

adjust the center eye in steps of $36 \text{ mV}_{\text{ppd}}$, and the upper and lower eyes in steps of $18 \text{ mV}_{\text{ppd}}$. This concept is shown in Fig. 15 for the case where the center eye height is reduced while the outer eye height is increased. Nominally, the equal level spacing is provided by a 2:1 ratio of MSB slices to LSB slices. An asymmetry between the outer eye height and the center eye height can be created by changing this ratio while maintaining the same termination impedance. For example, to decrease the +1 level by Δ (and also increase the -1 level by Δ), fewer slices are activated in the MSB cluster and more slices are activated in the LSB cluster. In general, this allows the eye to be pre-distorted to potentially compensate for any channel nonlinearity, such as those that arise in some optical applications. In terms of clock generation, the TX uses quarter-rate (QR) clocks for the 4:2 multiplexer (MUX) and an 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. DCC for the HR clock is implemented as shown in Fig. 16(b). In [4], DCD correction is implemented by adjusting pull-up and pull-down strengths of inverters in the clock path with tri-state inverters as shown in Fig. 16(a) where the tri-state inverters are in parallel with the main clock path. In this paper, the tri-state buffers are in a back-to-back configuration, reducing loading at the input since the parallel slices are no longer needed. This provides sharper edges and allows a smaller DCD adjustment resolution of 250 fs and a range of $\pm 1 \text{ ps}$. The simulated DCD varies around 70 fs across corners and the range is intended for mismatch control.

V. MEASUREMENTS

The prototype transceiver was fabricated in a TSMC 16-nm FinFET CMOS process. The die photograph and the RX-AFE layout floor plan are shown in Fig. 17. The TX occupies an area of $500 \mu\text{m} \times 180 \mu\text{m}$, while the RX-AFE occupies an area of $650 \mu\text{m} \times 250 \mu\text{m}$. For the TX, measurements using a 16-GHz clock pattern show a random jitter (RJ)

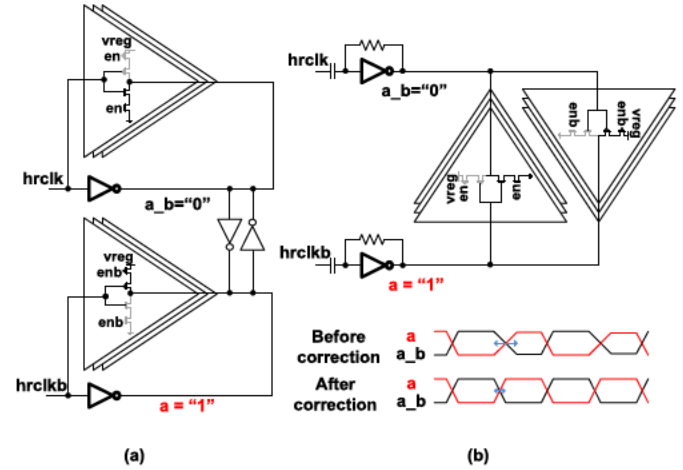


Fig. 16. TX DCC (a) prior art [4], and (b) this paper.

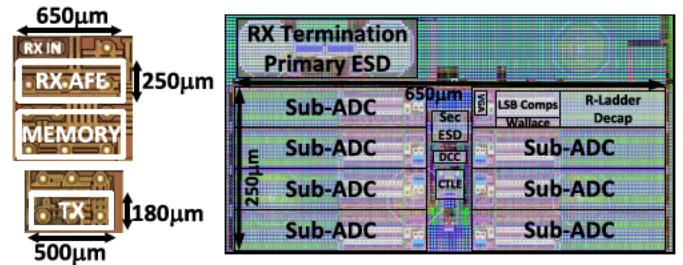
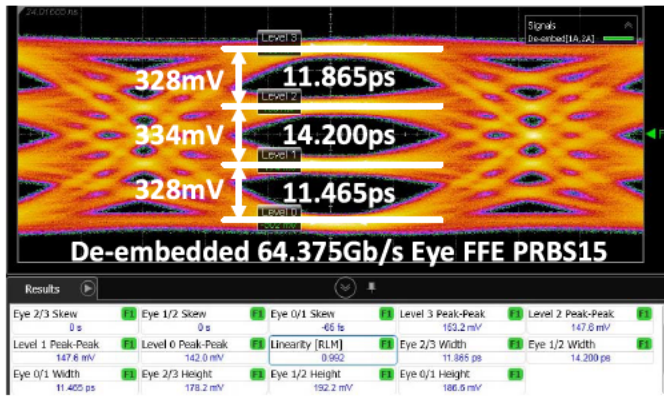


Fig. 17. Die photograph and layout floorplan of RX-AFE.

of $162\text{-fs}_{\text{rms}}$, and a total jitter (TJ) of 2.82 ps at a BER of $1e - 12$ [the clock is provided by an on-chip phase-locked loop (PLL) with a maximum frequency of 16.09375 GHz]. The de-embedded eye diagrams at 64.375 Gb/s with the TX FFE set to $[-1.5, 30, -1.5]$ and 4-PAM PRBS15 (single-bit stream, MSB bit followed by LSB bit mapped into a symbol) input are shown in Fig. 18. Fig. 18(a) shows the eye without pre-distortion, having a maximum horizontal opening of 14.2 ps and a maximum vertical opening of 334 mV. Level spacing corresponds to an level separation mismatch ratio (RLM) of 99.2% at 1 V_{ppd} full swing. With level pre-distortion active, Fig. 18(b) shows the middle eye height attenuated by 52 mV while the horizontal opening of the top and bottom eyes is increased slightly.

For the RX-AFE, the ADC is calibrated by adjusting the VGA gain to center the flash comparator calibration DAC codes (coarse gain mismatch calibration) and then an OFF-chip DAC (resolution 2 mV) is used to tune the comparator thresholds in each sub-ADC by applying the corresponding dc trip-point in turn, thus correcting for any static nonlinearity. After analog only correction, a residual gain and offset mismatch between the eight sub-ADCs of 2.2% and 2.18 LSB, respectively, are corrected off-chip using a single-tone low-frequency calibration step. Fig. 19(a) shows the SNDR/spurious-free dynamic range (SFDR) versus frequency for the RX-AFE at different input full-scale ranges (FSRs) and sampling frequencies. As an RX, a FSR of 400 and 500 mV_{ppd} are used. The spectrum [16 k · pt fast Fourier transform (FFT)] for a Nyquist frequency input



(a)



(b)

Fig. 18. TX eye diagram. (a) De-embedded 64.375-Gb/s eye diagram with FFE set to $[-1.5, 30, -1.5]$ for PRBS15 input and (b) with pre-distortion control active, attenuating middle eye height.

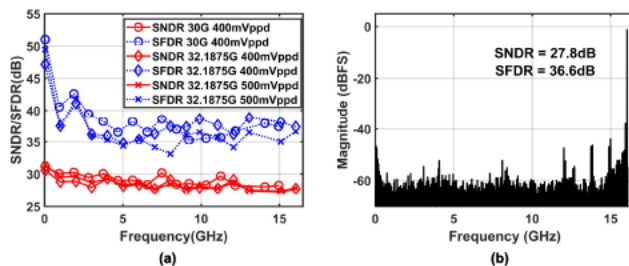


Fig. 19. RX-AFE frequency performance. (a) SNDR/SFDR versus frequency and (b) spectrum (16 k-pt FFT) for a 16-GHz Nyquist frequency input.

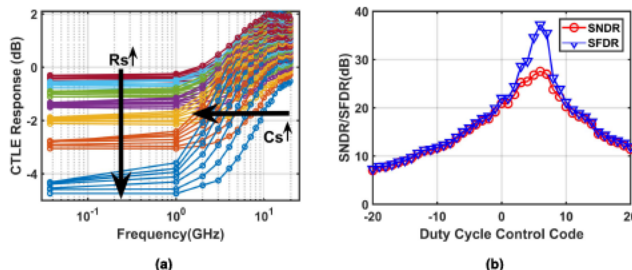


Fig. 20. (a) RX-AFE CTLE sampler response relative to flat setting. (b) SNDR/SFDR versus duty-cycle control code.

is shown in Fig. 19(b); the RX-AFE achieves an SNDR of 27.8 dB and SFDR of 36.6 dB.

Fig. 20(a) shows the performance of the CTLE sampler for all degeneration settings. The sampler exhibits a maximum boost of ~ 6 dB at the Nyquist frequency of 16 GHz with

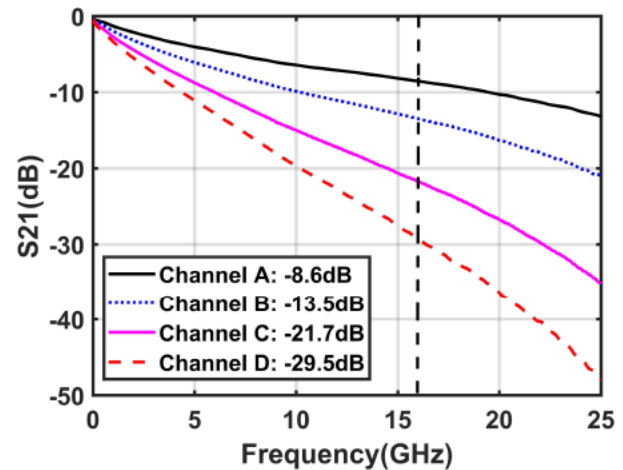


Fig. 21. Test link ILs: channels A–D ranging from 8.6- to 29.5-dB loss at the Nyquist frequency of 16 GHz.

the second pole over 20 GHz. The duty cycle of the sampler is corrected by sweeping the DCC code with a resolution of 250 fs to maximize SNDR for a Nyquist frequency input. Doing so, a DCD of 4.8% was observed and the skew tone is reduced to -45 dBFS by correction as shown in Fig. 20(b).

In order to perform link measurements, channels with varying IL at the Nyquist frequency were used as shown in Fig. 21. The shortest channel (IL = -8.6 dB), channel A, is a direct loop back consisting of a cable, the package and printed circuit board (PCB) loss on both the TX and RX side. Channels B, C, and D use an off-chip test PCB to increase the loss to 13.5, 21.7, and 29.5 dB, respectively. For channel A, the TX FFE is set to $[-13.6\%, 77.3\%, -9.1\%]$ and the CTLE is set to provide a boost of 5.6 dB with a dc gain of -1.5 dB. The RX-AFE is configured with a FSR of 500 mV_{ppd}. The eye diagram is open at the output of the ADC as shown in Fig. 22(a), indicating that the ADC can be configured as a slicer with only two comparators active per sub-ADC (due to folding). The ADC in slicer mode achieves a BER $< 1e-6$ while consuming 100 mW [excluding de-serializer, PLL, and clock and data recovery (CDR)]. For channel C, the eye is completely closed at the output of the ADC as shown in Fig. 22(b) even with the TX FFE set to $[-13.6\%, 63.7\%, -22.7\%]$ and CTLE set to a maximum boost of ~ 6 dB. An eight-tap FFE (two pre/five post-taps) in software is used to open the eye at a BER $< 1e-6$ with the ADC in 6-bit mode, which consumes 283.9 mW. For channel D with the highest loss, with the same TX FFE and CTLE settings, the FFE filter length is increased to 11 in order to open the eye at a BER $< 1e-4$. Since this work does not employ a CDR unit, bathtub curves, shown in Fig. 23, for channels A, C, and D are generated by finding and freezing the optimal TX FFE, CTLE, and DSP EQ coefficients at a chosen sampling phase and then sweeping the TX phase interpolator (PI) with a resolution of ~ 1 ps to cover the entire unit-interval (UI).

Channel B is used to investigate the greedy search-based power scaling. The TX FFE and RX CTLE are set to the optimal settings. The ADC is configured and calibrated for full

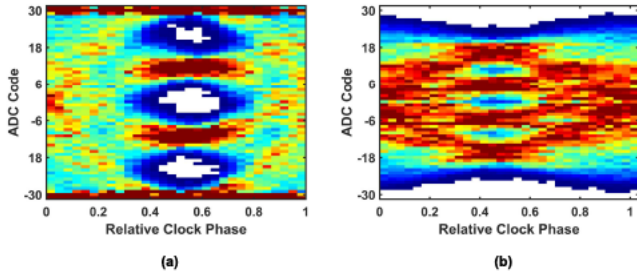


Fig. 22. ADC output eye diagrams (~ 52 -K samples) without DSP EQ, for (a) channel A (IL = -8.6 dB) showing an open eye and (b) channel C (IL = -21.7 dB) showing a closed eye, with TX FFE and CTLE at optimal settings.

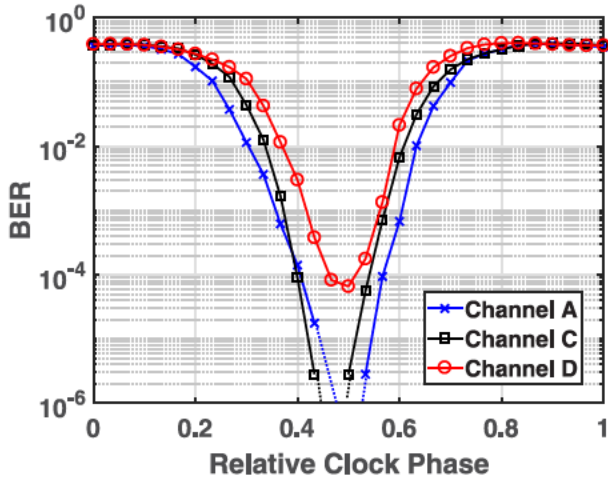


Fig. 23. Bathtub curves for channels A, C, and D generated by sweeping TX PI with all EQ coefficients frozen.

6-bit operation with a software six-tap FFE and power scaled for a target BER of $1e-5$ (to save testing time reading the ADC memory from the chip). Fig. 24 shows the experimental result with active comparators indicated by solid circles and inactive comparators indicated in white. After iteration 18, the BER starts increasing dramatically above 10^{-5} and the search can be terminated. The threshold levels at iteration 16 correspond to the same number of levels as a uniform 5-bit ADC and are used for comparison purposes. These levels are illustrated in Fig. 25 with the ADC output PDF (eye) plotted in conjunction. The BER obtained using the greedy search thresholds is almost a magnitude better $4.2e-6$ compared to $2.8e-5$ obtained using a 5-bit ADC configuration with uniformly spaced threshold levels. A link with a similar loss profile to channel B was then used to see the performance of the algorithm for various digital equalizer structures. Fig. 26 shows the 5-bit non-uniform levels chosen in the case of a six-tap FFE, one-tap DFE, and two-tap DFE following the ADC. For the DFE cases, the levels are more concentrated toward the center of the eye as expected. Clearly, the greedy search is able to co-optimize the equalizer architecture and the ADC quantization levels. In terms of convergence time, if the feedback loop is implemented on-chip, assuming the BER estimate requires $1e9$ bits and the equalizer

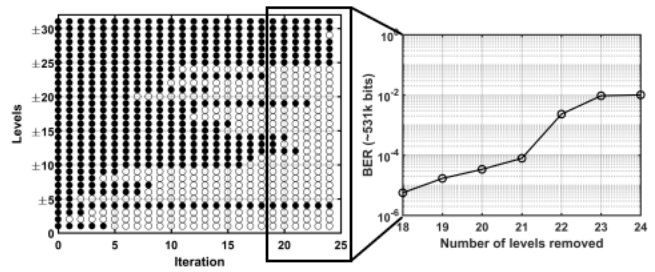


Fig. 24. Link power scaling with greedy search illustrating non-uniform quantization selection.

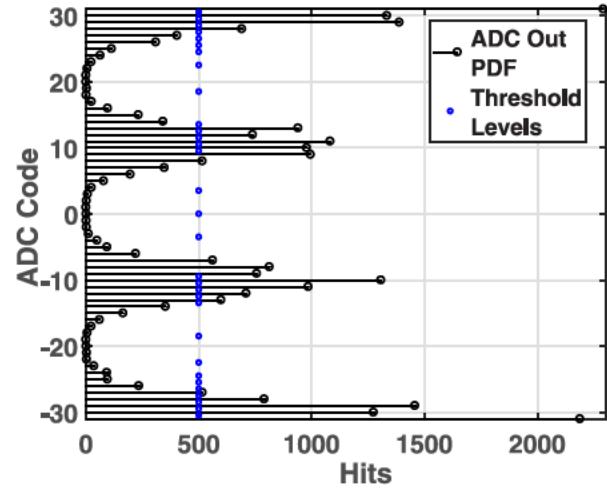


Fig. 25. Illustration of optimal 5-bit non-uniform levels: ADC output PDF (in black) at the optimal sampling point, greedy search levels shown in blue.

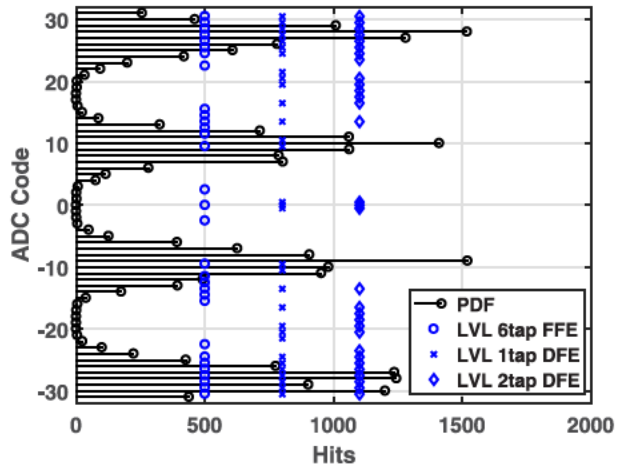


Fig. 26. Illustration of 5-bit non-uniform level selection using greedy search for different equalizer structures.

training requires $1e6$ bits, the total time for greedy search, in this case, would be approximately 6 s at this data rate.

Table II compares the performance of this transceiver to other 4-PAM transceivers operating at or above 50 Gb/s. The TX achieves the best power efficiency of 1.39 pJ/bit (89.7 mW) with the best RLM of 99% and RJ of 162-fs rms. The RX, in this paper, is the only one which enables non-uniform quantization-based power scaling at a data rate ≥ 50 Gb/s

TABLE II
SUMMARY AND COMPARISON OF ≥ 50 -Gb/s 4-PAM TX/RX

	[21]	[6]	[4]	[5]	This work
Data Rate (Gb/s)	64	50	56	56	64.375
RX ADC Res (bits)	-	7	8	7	6
RX AFE ENOB at Nyquist (bits)	-	4.9**	4.9	4.43	4.31
RX AFE Power (mW)	-	-	370*	215@7.4dB 245@32dB	100@8.6dB 283.9@29.5dB
TX Swing ($V_{pp-diff}$)	1.2	1.4	1.2	1	1.1
TX RJ ($f_{s,rms}$)	290	240	200	180	162
TX RLM (%)	94	-	97	98	99
TX Power (mW)	145	-	140	80.08	89.7
Supplies (V)	1.0/1.2	0.9/1.2	0.9/1.2	0.85/0.9/1.2	0.9/1.2
Link BER	-	2e-7 @35dB 2mV _{rms} ICN	<1e-8 @31dB 3.5mV _{rms} ICN	<1e-12 @32dB w/o XT	<1e-4 @29.5dB w/o XT
Link Power Efficiency (pJ/bit) (No DSP)	-	-	9.1	5.3@7.4dB 5.8@32dB	2.95@8.6dB 5.8@29.5dB
Chip Area (mm ²)	-	30.87***	2.8	8.81****	TX: 0.09 RX: 0.1625
Technology	28nm FDSOI	28nm	16nm FinFET	16nm FinFET	16nm FinFET

*Including retimer, **at 10GHz in figure

area including 2 RX/TX/PLL & DSP, *area including 4 RX/TX/PLL & DSP

with a novel greedy search approach. It allows for more power savings at low channel loss compared to other works; for instance [4], [5] achieves only 12.2% power reduction when link loss is decreased by more than 20 dB. In this paper, an RX-AFE power saving of 64.8% is obtained when the link loss is reduced by 21 dB. This power saving results from turning off 240 comparators ($\sim 720 \mu\text{W}/\text{comparator}$) and turning down the SF drive strength. The poorer BER performance, in this paper, can be attributed to several design choices: primarily the lower resolution of the quantizer, higher thermal noise, meta-stability of the LSB comparator and retimer at this data rate, and much simpler CTLE stage. For instance, the CTLE in [4] provides 14-dB boost at the Nyquist and due to the use of multiple stages, the channel shaping and noise performance can be optimized.

VI. CONCLUSION

A 64.375-Gb/s transceiver was demonstrated in TSMC 16-nm FinFET CMOS. The TX achieves an RLM of 99% at a power consumption of 1.39 pJ/bit. A pre-distortion mode improves the outer eye openings in the case of a compressive nonlinearity. The RX-AFE based on a 32-GS/s eight-way time-interleaved 6-bit ADC with a front-end two-way time-interleaved sampling CTLE achieves a SNDR of 27.8 dB at the Nyquist frequency with a power consumption of 283.9 mW. The RX-AFE scales its power consumption with link loss, offering a minimum power consumption of 100 mW in slicer mode. A greedy search is proposed to optimize the non-uniform quantization levels affording a graceful tradeoff between performance and power consumption.

ACKNOWLEDGMENT

The authors would like to thank Huawei for the logistic support, especially D. Dunwell, D. Cassan, and D. Tonietto,

and the Huawei layout team, especially M. A. Khan, D. Ilieva, M. Roberts, and T. Monson.

REFERENCES

- [1] *IEEE Standard for Ethernet—Amendment 10: Media Access Control Parameters, Physical Layers, and Management Parameters for 200 Gb/s and 400 Gb/s Operation*, IEEE Standard 802.3bs-2017, Dec. 2017, pp. 1–372.
- [2] Optical Internetworking Forum, “IA title: Common electrical I/O (CEI)—electrical and jitter interoperability agreements for 6G+ bps, 11G+ bps and 25G+ bps I/O,” I. Forum, Fremont, CA, USA, Tech. Rep. IA # OIF-CEI-04.0, 2017. [Online]. Available: <http://www.oiforum.com/wp-content/uploads/OIF-CEI-04.0.pdf>
- [3] J. Im *et al.*, “A 40-to-56Gb/s PAM-4 receiver with 10-tap direct decision-feedback equalization in 16nm FinFET,” in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2017, pp. 114–115.
- [4] Y. Frans *et al.*, “A 56-Gb/s PAM4 wireline transceiver using a 32-way time-interleaved SAR ADC in 16-nm FinFET,” *IEEE J. Solid-State Circuits*, vol. 52, no. 4, pp. 1101–1110, Apr. 2017.
- [5] P. Upadhyaya *et al.*, “A fully adaptive 19-to-56Gb/s PAM-4 wireline transceiver with a configurable ADC in 16nm FinFET,” in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2018, pp. 108–110.
- [6] K. Gopalakrishnan *et al.*, “A 40/50/100Gb/s PAM-4 Ethernet transceiver in 28nm CMOS,” in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2016, pp. 62–63.
- [7] Aurangozeb, A. D. Hossain, M. Mohammad, and M. Hossain, “Channel-adaptive ADC and TDC for 28 Gb/s PAM-4 digital receiver,” *IEEE J. Solid-State Circuits*, vol. 53, no. 3, pp. 772–788, Mar. 2018.
- [8] J. Kim *et al.*, “Equalizer design and performance trade-offs in ADC-based serial links,” *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 58, no. 9, pp. 2096–2107, Sep. 2011.
- [9] E.-H. Chen, R. Yousry, and C.-K. K. Yang, “Power optimized ADC-based serial link receiver,” *IEEE J. Solid-State Circuits*, vol. 47, no. 4, pp. 938–951, Apr. 2012.
- [10] A. D. Hossain, Aurangozeb, M. Mohammad, and M. Hossain, “A 35 mW 10 Gb/s ADC-DSP less direct digital sequence detector and equalizer in 65nm CMOS,” in *Proc. IEEE Symp. VLSI Circuits (VLSI-Circuits)*, Jun. 2016, pp. 1–2.
- [11] Y. Lin *et al.*, “A study of BER-optimal ADC-based receiver for serial links,” *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 63, no. 5, pp. 693–704, May 2016.
- [12] S. Lloyd, “Least squares quantization in PCM,” *IEEE Trans. Inf. Theory*, vol. IT-28, no. 2, pp. 129–137, Mar. 1982.
- [13] R. Narasimha, M. Lu, N. R. Shanbhag, and A. C. Singer, “BER-optimal analog-to-digital converters for communication links,” *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3683–3691, Jul. 2012.
- [14] IEEE 802.3cd Ethernet Task Group. (2018). *IEEE 802.3cd 50Gb/s, 100Gb/s, and 200 Gb/s Ethernet Task Force Contributed Channel Data*. [Online]. Available: <http://www.ieee802.org/3/cd/public/channel/index.html>
- [15] M. El-Chammas and B. Murmann, “General analysis on the impact of phase-skew in time-interleaved ADCs,” *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 56, no. 5, pp. 902–910, May 2009.
- [16] Y. Duan and E. Alon, “A 12.8 GS/s time-interleaved ADC with 25 GHz effective resolution bandwidth and 4.6 ENOB,” *IEEE J. Solid-State Circuits*, vol. 49, no. 8, pp. 1725–1738, Aug. 2014.
- [17] L. Wang, M. LaCroix, and A. C. Carusone, “A 4-GS/s single channel reconfigurable folding flash ADC for wireline applications in 16-nm FinFET,” *IEEE Trans. Circuits Syst., II, Exp. Briefs*, vol. 64, no. 12, pp. 1367–1371, Dec. 2017.
- [18] B. Zhang *et al.*, “A 40 nm CMOS 195 mW/55 mW dual-path receiver AFE for multi-standard 8.5–11.5 Gb/s serial links,” *IEEE J. Solid-State Circuits*, vol. 50, no. 2, pp. 426–439, Feb. 2015.
- [19] M. Miyahara and A. Matsuzawa, “A low-offset latched comparator using zero-static power dynamic offset cancellation technique,” in *Proc. IEEE Asian Solid State Circuits Conf. (ASSCC)*, Nov. 2009, pp. 233–236.
- [20] K.-M. Lei, P.-I. Mak, and R. P. Martins, “Systematic analysis and cancellation of kickback noise in a dynamic latched comparator,” *Analog Integr. Circuits Signal Process.*, vol. 77, no. 2, pp. 277–284, Nov. 2013.
- [21] G. Steffan *et al.*, “A 64Gb/s PAM-4 transmitter with 4-Tap FFE and 2.26pJ/b energy efficiency in 28nm CMOS FDSOI,” in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2017, pp. 116–117.



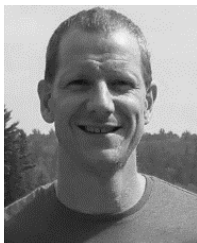
Luke Wang (S'18) received the B.A.Sc. degree in electrical engineering from the University of Waterloo, Waterloo, ON, Canada, in 2011, and the M.A.Sc. degree from the University of Toronto, Toronto, ON, Canada, in 2014, where he is currently pursuing the Ph.D. degree, with a focus on optimizing reconfigurable analog-to-digital converters for wireline applications.

In 2013, he joined Broadcom Corporation, Irvine, CA, USA, as an Intern, working on a 28-nm SerDes project. From 2016 to 2017, he was an Intern with Huawei Canada, Toronto, ON, USA, where he was involved in investigations of analog-to-digital converter (ADC)-based receiver architectures for 4-pulse-amplitude modulation (PAM) short reach to median-reach electrical SerDes links.



Yingying Fu received the B.A.Sc. and M.A.Sc. degrees in electrical engineering from the University of Toronto, Toronto, ON, Canada, in 2012 and 2015, respectively.

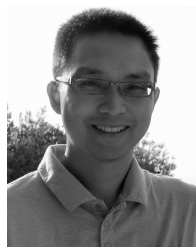
In 2015, she joined Huawei Canada, Markham, ON, Canada, where she is working on high-speed SerDes circuit design.



Marc-Andre LaCroix (S'97–M'02) received the B.Sc.Eng degree in electrical engineering from the University of New Brunswick, Fredericton, NB, Canada, in 2000, and the M.A.Sc degree in electrical engineering from Carleton University, Ottawa, ON, Canada, in 2002.

He was with Nortel Networks, ON, Canada, STMicroelectronics, ON, Canada, and Gennum Corporation, ON, Canada. Since 2011, he has been with Huawei Technologies Co, Kanata, ON, Canada. He is currently with Huawei as a

Distinguished Engineer and has served as a Design Leader and Architect for several generations of high-speed wireline transceivers. He is currently involved in the development of the next generation of transceivers targeting line rates of 112-200G.



Euhan Chong (M'13) received the B.A.Sc. degree in systems design engineering from the University of Waterloo, Waterloo, ON, Canada, in 2001, and the M.A.Sc. degree in electrical engineering from the University of Toronto, Toronto, ON, in 2004.

From 2004 to 2011, he was with Synopsys, Mississauga, ON, Canada, and then with Semtech, Ottawa, ON, Canada. Since 2011, he has been with Huawei Canada, Kanata, ON, Canada, where he is working on mixed-signal circuits for high-speed serial links.



Anthony Chan Carusone (S'96–M'02–SM'08) received the Ph.D. degree from the University of Toronto, Toronto, ON, Canada, in 2002.

Since 2002, he has been with the Department of Electrical and Computer Engineering, University of Toronto, where he is currently a Professor. He is also an occasional consultant to Industry in the areas of integrated circuit design and digital communication. He has co-authored the Best Student Papers at the 2007, 2008, and 2011 Custom Integrated Circuits Conferences, the Best Invited Paper at the 2010 Custom Integrated Circuits Conference, the Best Paper at the 2005 Compound Semiconductor Integrated Circuits Symposium, and the Best Young Scientist Paper at the 2014 European Solid-State Circuits Conference. He has also co-authored, along with D. Johns and K. Martin, the 2nd edition of the textbook "*Analog Integrated Circuit Design*."

Dr. Chan Carusone currently serves as a member for the Technical Program Committee of the International Solid-State Circuits Conference. He was an Associate Editor of the IEEE *Journal of Solid-State Circuits* from 2010 to 2017. He was the Editor-in-Chief of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II: EXPRESS BRIEFS IN 2009. He has served on the Technical Program Committees for the Custom Integrated Circuits Conference and the VLSI Circuits Symposium. He was a Distinguished Lecturer of the IEEE Solid-State Circuits Society from 2015 to 2017.