An 80 mW 40 Gb/s 7-Tap *T*/2-Spaced Feed-Forward Equalizer in 65 nm CMOS

Afshin Momtaz, *Member, IEEE*, and Michael M. Green, *Member, IEEE*

***Abstract*—A 7-tap 40 Gb/s FFE using a 65 nm standard CMOS process is described. A number of broadbanding and calibration techniques are used, which allow high-speed operation while consuming 80 mW from a 1 V supply. ESD protection is added to 40 Gb/s IOs and an inexpensive plastic package is used to make the chip closer to a commercial product. The measured tap delay frequency response variation is less than 1 dB up to 20 GHz and tap-to-tap delay variation is less than 0.3 ps. More than 50% vertical and 70% horizontal eye opening from a closed input eye are observed. The use of a CMOS process enables further integration of this core into a DFE equalizer or a CDR/Demux based receiver.**

***Index Terms*—CMOS analog integrated circuits, current mode logic, FFE, broadband communication, equalizers.**

I. Introduction

OPTICAL communication systems have been used for high-speed data transmission since the early 1970's. To satisfy the demand for greater network capacity, the data rate of current broadband systems has been pushed to 10 and 40 Gb/s. At these data rates, it is no longer possible to neglect the bandwidth limitations of the channel. Dispersed isolated pulses interfere with each other leading to eye diagram closure and an increase in bit error rate (BER) at the receiver. At the 40 Gb/s rate, deployment of dispersion compensation or equalization is necessary. Due to its fast adaptation speed and ease of integration within the transceiver, electronic dispersion compensation (EDC) is receiving a great deal of attention. A feed-forward equalizer (FFE) is currently the most practical implementation of EDC for 40 Gb/s data rates, reflecting its advantages as a simple structure with moderate design complexity.

An FFE can generate a wide variety of different linear transfer functions, making it useful for electrical and optical channel impairment mitigation or signal waveform optimization [1], [2]. The block diagram of an FFE is shown in Fig. 1, and its input/output relationship is given by

$$y(nT) = \sum_{i=1}^{M} C_i \cdot x((n+1-i)T) \tag{1}$$

Manuscript received July 22, 2009; revised October 26, 2009. Current version published February 24, 2010. This paper was approved by Associate Editor Jafar Savoj. This work was supported by Broadcom Corporation during chip manufacturing and testing.

A. Momtaz is with Broadcom Corporation, Irvine, CA 92618 USA.

M. M. Green is with the Department of Electrical Engineering and Computer Science, University of California, Irvine, CA 92697-2625 USA (e-mail: mgreen@uci.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSSC.2009.2039268

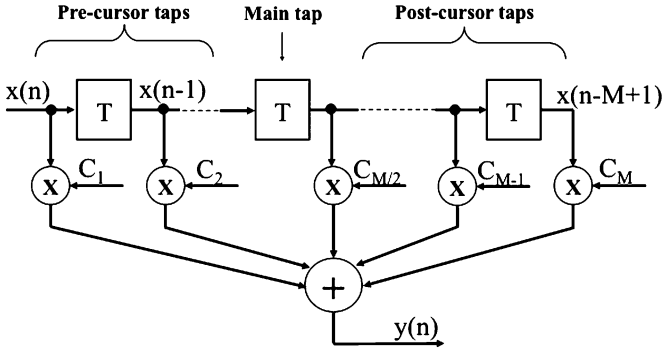

Fig. 1. M-tap FFE block diagram.

where $x(n)$ and $y(n)$ are the input and output signals respectively; C_i is the ith coefficient; and M is the number of taps. The input signal, $x(n)$, propagates along a delay line composed of M unit-interval delay elements. The delayed signals are then multiplied by adjustable coefficients and finally summed together. One of the taps near the center is commonly referred to as the main tap; the taps that follow (precede) the main tap are called the post-cursor (pre-cursor) taps.

The FFE equalization capability can be examined in the frequency domain. The transfer function of a 7-tap FFE with $T = 25$ ps is plotted under different conditions in Fig. 2, where C_4 represents the middle tap and C_1 to C_3 are the pre-cursor taps. In these plots, C_4 is held at unity while each tap is varied one at a time with the other taps set to zero. As shown in Fig. 2(a), by varying C_1 from 0.5 to -0.5, frequencies near 6.6 GHz can amplified or attenuated. As shown in Fig. 2(b), varying C_2 has a similar effect at frequencies near 10 GHz. As shown in Fig. 2(c), varying C_3 affects the peaking near 20 GHz. This behavior can be easily understood by realizing that C_3, C_2, and C_1 are one, two, and three taps away from the main tap C_4, respectively. Thus, their frequency responses differ only by the appropriate frequency-scaling ratio. By combining different tap values, a wide variety of filter transfer functions can be created. This flexibility in changing various aspects of the filter characteristics is the main advantage of FFE over peaking-type continuous-time equalizers (e.g., [3], [4]).

The system performance of an FFE is dictated primarily by two parameters: the tap spacing (also known as tap delay) and the number of taps. Fractionally-spaced FFE structures have been utilized for more than two decades. In particular, Gitlin's work on $T/2$ spaced equalization [5] demonstrates that this type of equalizer not only reduces aliasing but also directly improves performance. Such a $T/2$-spaced structure doubles the equalizer frequency domain range, as illustrated in Fig. 3. In this figure,

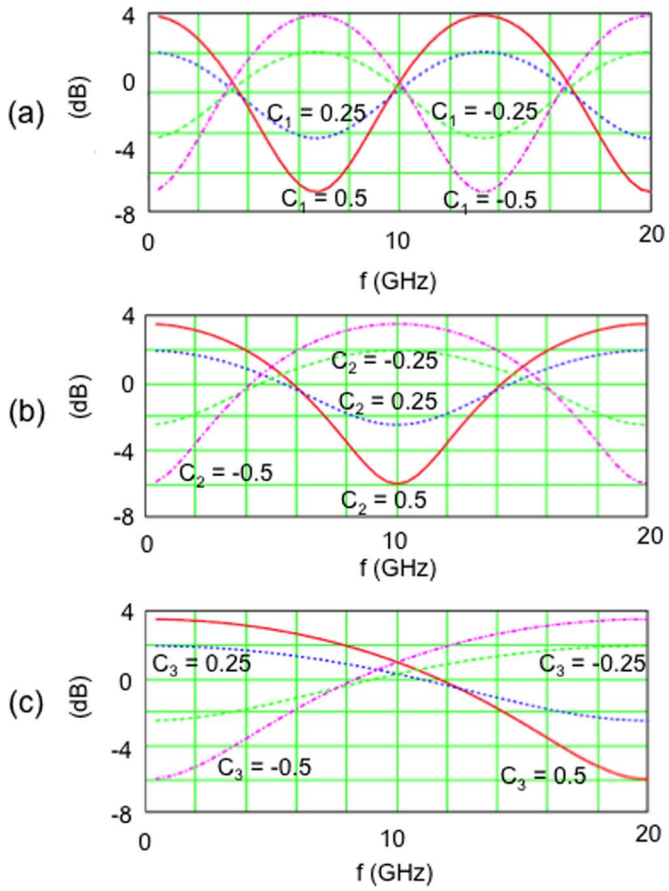

Fig. 2. 7-tap FFE transfer function as only one tap is modified: (a) C_1 is changed; (b) C_2 is changed; (c) C_3 is changed.

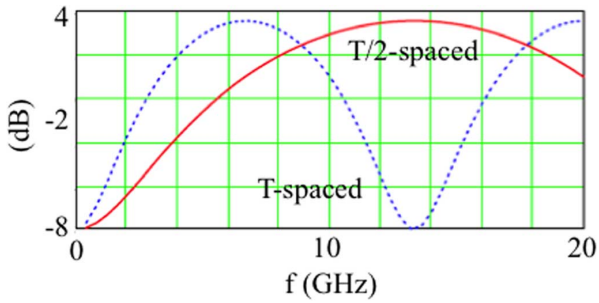

Fig. 3. Comparison between transfer function of a 7-tap T-spaced and $T/2$-spaced FFE.

the transfer functions of a T-spaced and a $T/2$-spaced FFE, each with 7 taps, are compared with all coefficients set to zero except for $C_4 = 1$ and $C_1 = -0.5$. Although both transfer functions have identical shape, a 2X frequency scaling can be observed in the case of $T/2$-spaced equalizer. By reducing the tap delay further (for example to $T/3$), better equalization of high-frequency components can be achieved (3X frequency scaling) at the price of less total ISI span compensation (1/3rd total tap delay), for a given number of taps.

In addition, system-level simulations of an SMF link and a $T/2$-spaced FFE [5] show that if the number of taps is increased beyond 7, the performance improvement is marginal. This limitation is directly related to the ISI from the SMF pulse response. Once implementation non-idealities and power consumption of additional taps are also considered, a 7-tap $T/2$-spaced structure becomes a reasonable compromise.

Compared to III–V technologies or bipolar processes, the scalability, availability, ease of integration and lower static power consumption of a CMOS process makes it desirable for implementing the FFE. However, the lower process speed and lower power supply voltage of CMOS create challenging obstacles for implementing a 40 Gbps FFE with as many as 7 taps. In this work, first presented in [7], through various architectural and circuit design techniques, these difficulties are addressed.

The remainder of this paper is organized as follows. Section II describes the circuit implementation of various blocks. The measured results are presented in Section III. Finally, Section IV concludes this paper.

II. Circuit Design

A. Architecture

The architecture of previously published 40 Gb/s feed-forward equalizers (e.g., [1], [8]), is shown in Fig. 4(a). Each tap delay is implemented through two separate delay elements: one at the input of the multipliers and one at the output, with the overall delay being the sum of the two individual delays. Variable transconductance cells perform the multiplication and then the delayed versions of their current outputs are summed together and converted back to a voltage through a termination resistor R_s. Here, the FFE input and output signals are located on the same side of the block, and in the same vicinity. If the chip contains only the equalizer (as is the case in this design), the input and output signals travel closely not only on the die but also on the package and the board. The coupling between them can cause severe signal integrity issues, degrading the equalizer performance. In addition, when multiple high-speed blocks are cascaded, their interconnect length is minimized when the input and output of each block are located on opposite sides. For example, if the FFE were followed by a CDR, the interconnect between the two blocks would be long due to the FFE input and output being located on the same side, leading to sub-optimum performance.

In the proposed architecture shown in Fig. 4(b), the input and output are naturally located on opposite sides of the equalizer, minimizing their coupling and simplifying the connection to the input of the next block. Another difference is that the overall tap delay is now given by the *difference* between the two individual delay elements. This delay subtraction minimizes the equalizer dependence on the passive delay element modeling as will be discussed in the next section.

B. Delay Element

A delay element can be implemented by using either a passive transmission line or an active unity-gain buffer. On-chip transmission lines have been used in various FFEs [1], [8] with low power dissipation being their main advantage over active unity-gain buffers. Transmission lines can be formed by strip

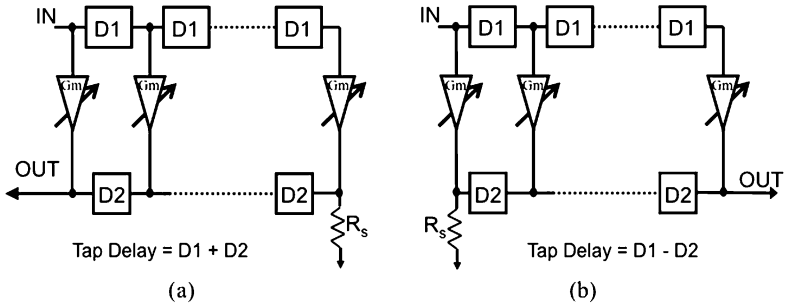

Fig. 4. FFE architectures: (a) conventional approach; (b) proposed approach.

lines, coplanar waveguides or lumped elements. In a lumped element transmission line, on-chip spiral inductors and capacitors are cascaded. At frequencies above 10 Gb/s, the parasitic capacitances of the transistors used in the FFE multipliers usually play the role of the transmission line capacitors. Thus, in contrast to the active delay approach where multiplier input and output capacitances directly limit the delay element bandwidth, the lumped element topology absorbs the capacitance and therefore reduces the bandwidth degradation. The lumped element approach, however, does suffer from some disadvantages. First, because multiple inductors are connected in series, the accuracy of their models is critical in predicting the FFE behavior. Second, the parasitic resistance of the inductors and their interconnections accumulates and limits the number of realizable FFE taps and the total delay of all taps. Third, the gain/loss of each tap is not well controlled, which reduces the overall equalizer performance. Due to this limitation, the highest total delay of all taps of 40 Gb/s FFE published to date has not exceeded 75 ps [9]–[11], [15].

We propose a solution that combines both approaches and generates the required tap delay through the use of both passive and active delay elements. Fig. 5 illustrates the FFE tap delay realization where active elements are used at the multiplier inputs, and passive elements are placed at the outputs. The active elements isolate each tap, and eliminate the need for larger die area for transmission lines. In addition, because transmission line modeling is not supported by industry standard CMOS CAD tools, the use of active elements is also more attractive from a practical point of view. At the same time the output currents are delayed through passive elements, absorbing the multiplier output capacitance and providing large bandwidth at the output. In this structure, the effective tap delay is the difference between the two delay elements. The delays for the active and passive elements are designed to be 15.5 ps and 3 ps, respectively, resulting in an effective tap delay of 12.5 ps. In addition, since the passive delay element accounts for only 25% of the total tap delay, its modeling inaccuracy plays a smaller role in the equalizer performance.

1) Active Delay Path: Fig. 6 shows the active delay element structure used in this design. Various techniques, described as follows, have been used to overcome conventional active delay element shortcomings mentioned previously.

Gain Control: The delay cell gain should be close to unity; any variation needs to be compensated by adjusting the FFE tap coefficients, which results in smaller available tap range and

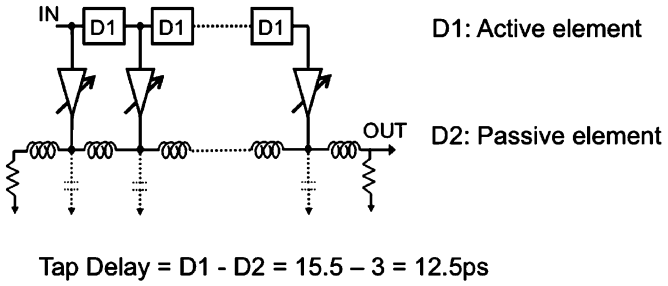

Fig. 5. The proposed tap delay implementation.

hence diminished equalization power. Thus, it is critical that the gain variation across PVT is minimized. The simplified expression for the gain is given as

$$A = g_m \cdot R_1 \tag{2}$$

where g_m is the transconductance of transistors M_1 and M_2 and transistor output impedances are ignored since they are much larger than R_1. By using a constant-gm biasing scheme [12], we can arrange to have $g_m = M/R_B$, where R_B is the value of a resistor used in the biasing circuitry and M is a process-independent constant given by

$$M = \sqrt{\frac{2 \cdot (W/L)_{1,2}}{(W/L)_B}} \cdot \left(1 - \frac{1}{\sqrt{N}}\right) \tag{3}$$

where $(W/L)_{1,2}$ corresponds to transistors M_1 and $M_2, (W/L)_B$ corresponds to the bias transistor, and N is the transistor size ratio used in the bias block [12]. Hence, we can write

$$A = M \cdot \frac{R_1}{R_B}. \tag{4}$$

Because M and R_1/R_B are PVT-independent constants, the dc gain can be well controlled. The validity of (4) is contingent on having all transistors biased in the saturation region. Thus, the transistors are appropriately sized and level-shifting resistor R_2 is added (Fig. 6). Use of these biasing techniques minimizes the dc gain variation to less than ± 0.4 dB across all PVT corners.

Bandwidth Enhancement: The design of the Fig. 6 delay element begins with a unity-gain CML buffer. By optimizing the differential pair sizes, load resistance, and tail current, a bandwidth of 11 GHz can be obtained using a standard CML buffer in the 65 nm CMOS process. Next, shunt-peaking inductors L_2

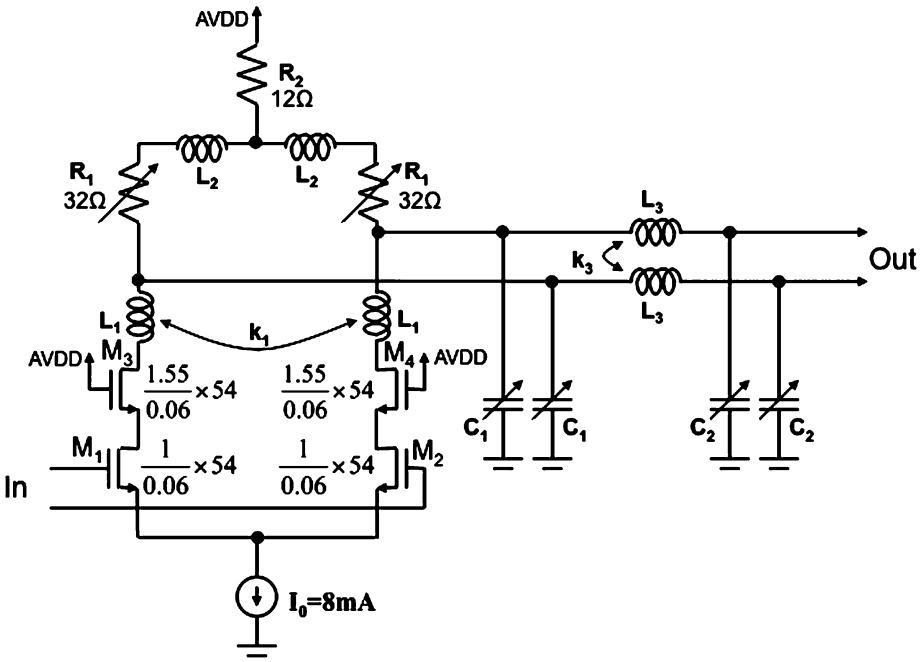

Fig. 6. Active delay cell transistor level schematic.

are added and the buffer parameters are re-optimized, increasing the bandwidth to 19 GHz. The addition of cascode transistors M_3 and M_4 reduces the Miller effect and the effective input capacitance due to $C_{\rm gd}$ by 20%, further increasing the bandwidth to 21 GHz. The introduction of cascode transistors has another benefit: When multiple delay cells are cascaded, the load of one can be seen by the previous one through the $C_{\rm gd}$ of transistors M_1 and M_2. At multi-GHz frequencies, the admittance of this capacitor is large enough that this multi-stage interaction becomes significant. The cascode transistors M_3 and M_4 reduce this interaction by improving the isolation between the delay cell input and output nodes. Finally double series-peaking, implemented by L_1 and L_3, is added to the delay cell, pushing the bandwidth to 41 GHz by allowing the various capacitors in the circuit to charge one at a time rather than in parallel. Since series peaking in general increases bandwidth while also increasing delay, a relatively large delay can easily be realized while maintaining a high bandwidth. For this design the nominal delay time is set to 15.5 ps. Fig. 7 summarizes the simulation bandwidth data for the above cases and shows the benefit of each added technique.

Time Delay Control: In the 65 nm CMOS process, P+ Poly resistors can vary up to $\pm 12\%$. Since the CML load resistance directly impacts CML stage time delay, the calibration of the resistors will minimize the delay variations. Fig. 8 shows the detail of the load calibration circuitry, composed of parallel branches of poly resistors in series with pMOS switches. A 3-bit binary-coded digital signal is used to control the effective impedance of the load; as the code increases from 000 to 111, the effective load resistance is uniformly increased.

The M_1 width and the R_1 value are specifically chosen so that the total resistance of their branch is 10 times larger than R_0. Similarly, the M_2 size and the R_2 value are chosen to make sure their branch resistance is 20 times larger than R_0, and the

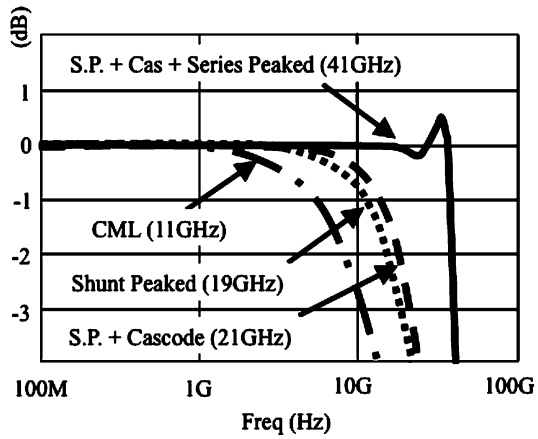

Fig. 7. Active delay element bandwidth enhancement through different broadbanding techniques.

resistance of the R_3 branch is 40 times larger than R_0. This resistance combination realizes step sizes each equal to 2.5%.

By comparing an on-chip resistor with an off-chip one, the correct value for the 3-bit control signal could be determined. Although this resistance comparator is not included on this chip, its implementation has been reported elsewhere [13]. The addition of resistor calibration reduces the time delay variation due to the PVT from 6.5 ps to less than 2.5 ps. Although by increasing the resolution of resistor calibration beyond 3 bits finer delay variation could be achieved, the improvement would be relatively small compared to the additional required circuitry and complexity.

By using a similar calibrated resistor in the biasing block, the term R_1/R_B is kept process independent and hence the active delay element gain does not change as the resistors are calibrated as can be seen from (4).

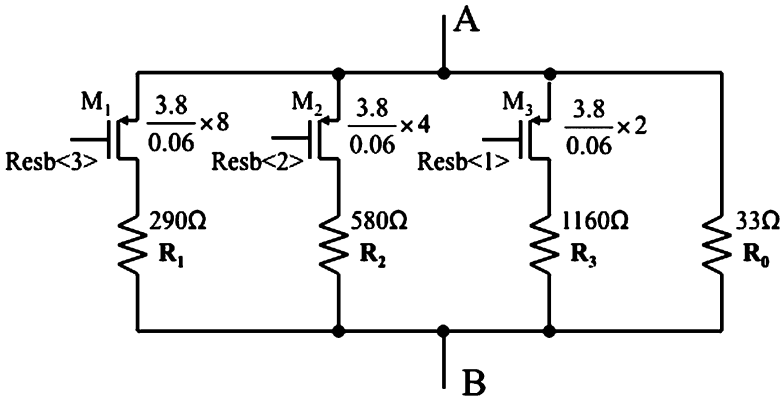

Fig. 8. Active delay element calibrated load resistor.

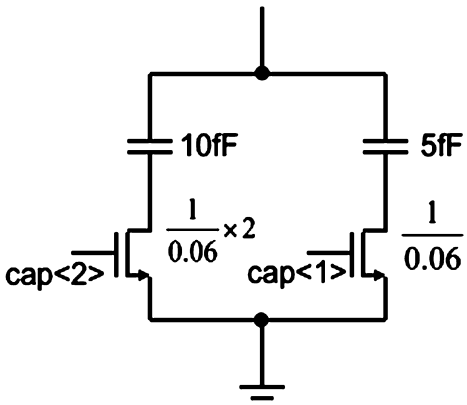

Fig. 9. Active delay element switchable capacitors used for temperature compensation.

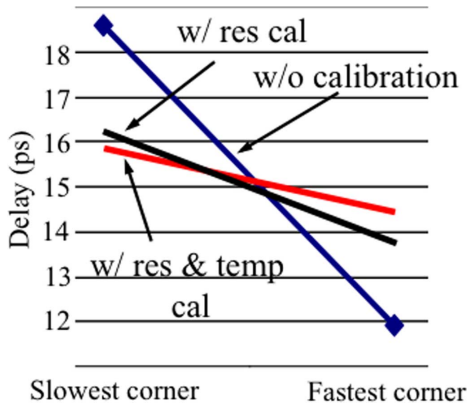

Fig. 10. Effect of different techniques on delay variation across PVT.

Moreover, the delay through the active element is temperature sensitive; at high temperature, the delay increases as the transistors slow down. To compensate for this effect, capacitors C_1 and C_2 in Fig. 6 are each realized as shown in Fig. 9. By adjusting these capacitances, the delay can be modified to cancel out the change due to temperature. Digital signals, each with 2 bits resolution, control the values of C_1 and C_2, providing 8 different settings where each is used for a specific temperature range. As a result, the time delay variation is further reduced to less than 1.5 ps. Fig. 10 illustrates the benefit of resistor and temperature calibration on the time delay variation.

Unlike resistor calibration which is performed only at the chip power up, capacitors C_1 and C_2 need to be adjusted as the temperature changes. To avoid glitches in the data path, thermometer-based implementation is required ensuring that only one capacitor is turned ON or OFF at a given time. Because the temperature adaptation loop is not fully integrated in this chip, the simpler binary approach has been used (Fig. 9). The full adaption loop could be implemented, for example, by using the temperature sensitivity of the voltage across an on-chip diode. The diode and three bandgap-based reference voltages could be applied to a 2-bit ADC, generating the 2-bit control codes.

2) Passive Delay Path: As previously mentioned, the passive delay elements are used at the multiplier outputs. As illustrated in Fig. 11, the lumped-element approach is implemented where the parasitic capacitance C_m of the multiplier output nodes and on-chip spiral inductors L_m form the required capacitors and inductors. To maximize bandwidth, the transmission line characteristic impedance, Z_0, should be matched with the termination impedance of the summer R_s, i.e.,

$$Z_0 = \sqrt{\frac{L_m}{C_m}} = R_s. \tag{5}$$

On the other hand, the passive element time delay per section t_0 is a function of both L_m and C_m:

$$t_0 = \sqrt{L_m \cdot C_m}. \tag{6}$$

Combining (5) and (6), the required termination resistance can be calculated in terms of desired time delay t_0 and the multiplier output capacitance:

$$R_s = t_0/C_m. \tag{7}$$

Based on the multiplier design (described in the next section), C_m is 75 fF. From the architectural specification, the desired time delay is 3 ps. Using (5) and (7), R_s and L_m are calculated to be 45 Ω and 150 pH, respectively.

The input and output capacitances of the delay line are denoted as C_s and C_d, respectively, in Fig. 11. Using (6), the value of required inductance for L_d and L_s are calculated to be 125 pH and 85 pH, respectively. The values of L_m, L_d, and L_s are then fine tuned through ac simulations of the entire FFE. In particular, these values are selected to maximize the FFE bandwidth while

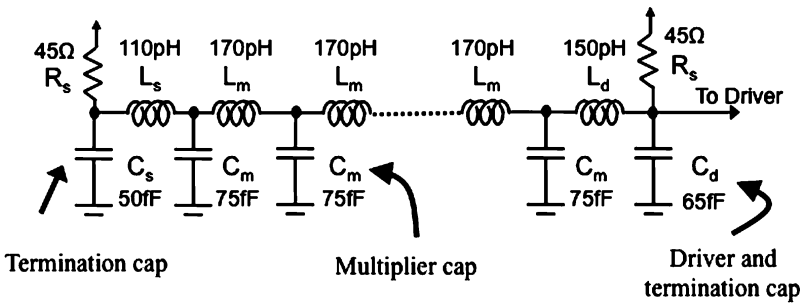

Fig. 11. Passive delay path implementation.

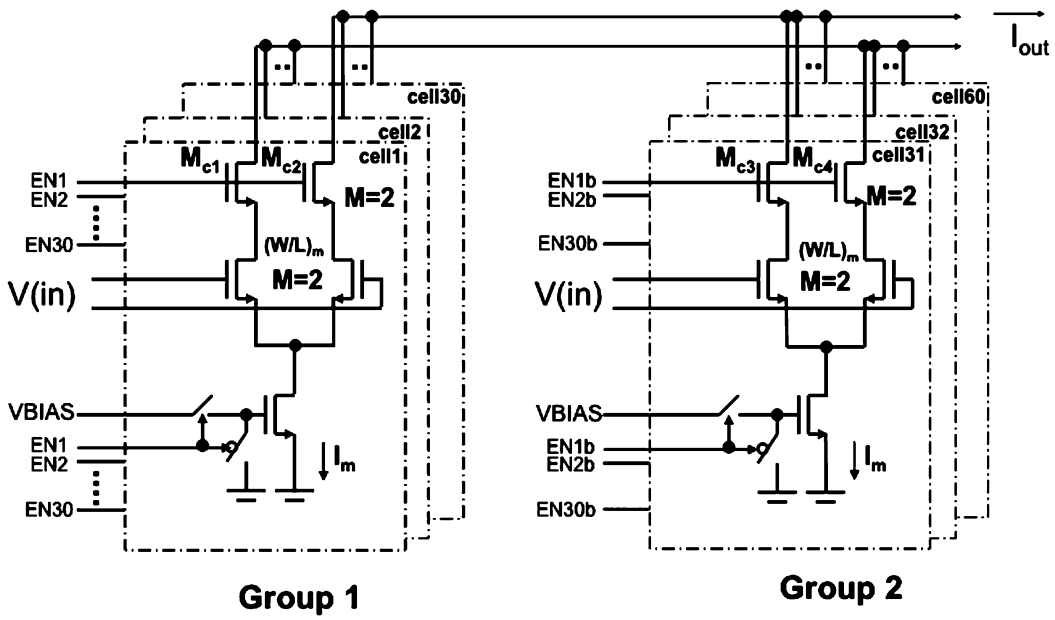

Fig. 12. The new multiplier transistor level diagram.

keeping the passband peaking below 1 dB. The final values of the passive delay path inductors are shown in Fig. 11.

It should be noted that capacitors C_m, C_d, and C_s represent the total capacitance at their corresponding nodes, including both the device and the interconnect capacitances. In addition, to minimize the effect of the process resistance variation on the summer performance, resistor R_s is implemented using the calibrated structure discussed earlier.

C. Multiplier Circuit

Implementation of the required multipliers can be challenging when linearity is crucial. Gilbert cells are commonly used [9], [14], but they suffer from a few shortcomings. First, although degeneration resistors can be added to increase the linear region of the multiplier, the multiplication step size still exhibits non-uniformity, especially at maximum and minimum gain settings. In addition, when the Gilbert cell is close to its gain limit, one of the high-speed differential pairs receives only a small amount of current, creating significant distortion at the output. Since an FFE is a linear filter, it is incapable of compensating nonlinearity; thus the distortion greatly degrades the equalizer performance. Finally, even if the current of one of the high-speed differential pairs is fully shut off, the transistors of this differential pair still have an impact on the multiplier gain. Specifically, these transistors conduct and behave as an equivalent resistor between the positive and negative output terminals. This resistance reduces the multiplier gain. Additional current or transistor size increase is required to compensate for the gain loss.

A new digitally controlled multiplier structure achieves uniform multiplication step sizes, lower distortion and higher maximum gain, while consuming the same power and area as the Gilbert cell topology. In this scheme, the gain is controlled by adjusting both the multiplier current and the differential pair W/L. As shown in Fig. 12, 60 identical transconductance unit cells form the multiplier. Each cell can be turned on or off through a digital control signal. Although all inputs to the 60 cells are connected together, the outputs are connected in two groups of 30 cells, each with opposite polarity. As a result, the first 30 transconductance cells (indicated as Group 1 in Fig. 12) add to the multiplier output current and the next 30 cells (indicated as Group 2) subtract from it. The gain is increased by turning on more cells from Group 1 and fewer cells from Group 2. To maintain a constant output common mode voltage, exactly 30 cells are on and the remaining 30 cells are off for any gain setting—that is, for each cell in Group 1 that is turned on, a corresponding cell in Group 2 is turned off. A 5-bit digital signal is decoded to 30 signals, each controlling one cell from Group 1 and one cell from Group 2.

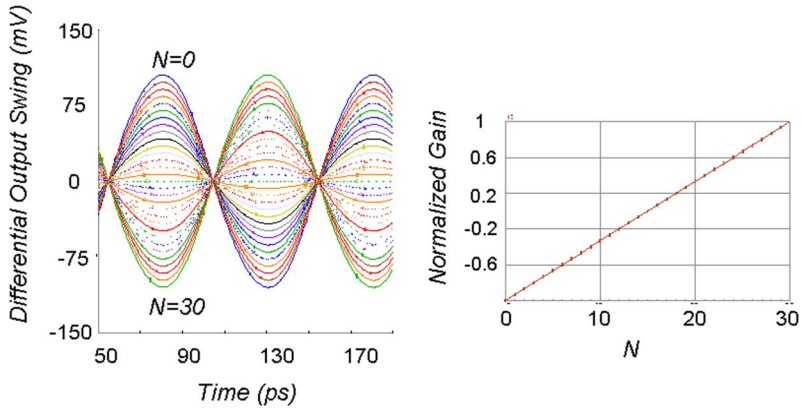

Fig. 13. The simulated results showing the multiplier behavior as a function of N: (a) the multiplier output waveform; (b) the normalized multiplier gain.

If N denotes the number of cells turned on in Group 1, then the multiplier gain is given by

$$A_{\text{mult}} = (2N - 30)R_s \cdot \sqrt{4 \cdot k_n' \cdot I_m \cdot \left(\frac{W}{L}\right)}. \tag{8}$$

Equation (8) indicates that the multiplier gain is now a *linear* function of the digital control signal value N; thus all the gain steps have uniform value, given by

$$\Delta A_{\text{mult}} = 4R_s \cdot \sqrt{k_n' \cdot I_m \cdot \left(\frac{W}{L}\right)}. \tag{9}$$

Fig. 13 shows the simulated multiplier behavior as the value of N is varied from 0 to 30. Specifically, the multiplier output swing in response to a 10 GHz input sine wave is shown in Fig. 13(a) and the normalized multiplier gain is shown in Fig. 13(b). The uniformity of the multiplier gain step is evident in these simulation results. The maximum gain can also be calculated from (8) by setting $N = 30$:

$$A_{\text{mult(max)}} = 60R_s \cdot \sqrt{k_n' \cdot I_m \cdot \left(\frac{W}{L}\right)}. \tag{10}$$

It can be shown that this expression is identical to the ideal maximum gain of a Gillbert cell using the same total current and input differential pair size. Furthermore, because the multiplier current and aspect ratio are both changed equally when the gain is modified, the V_{Dsat} of the differential pair transistors is kept constant. As a result, the output total harmonic distortion is not increased as the multiplier approaches its gain limits.

The cascode transistors M_{C1}–M_{C4} in Fig. 12 have dual purposes. When the unit cell is on, these transistors function as standard cascodes, reducing the Miller capacitance and providing isolation between the input and output loads. On the other hand, when the unit cell is off, M_{C1} to M_{C4} act as switches that have been shut off. As a result, the multiplier output conductance is reduced and the gain is increased. The multiplier gain is amplified without conducting any extra current or degrading FFE bandwidth.

It should be noted that although increasing the multiplier resolution beyond 5 bits provides finer FFE coefficient adjustment, the increase in multiplier size and parasitic capacitance reduces the equalizer overall bandwidth, producing no significant improvement in the chip equalization capability. Finally, the thermometer-based structure of the digital multiplier allows the FFE coefficients to be changed without introducing any glitches in the data path.

D. Summer and Tap Scaling

If all 7 taps of the FFE conducted identical currents, the total current at the summation node would be quite large, posing a number of challenges. First, because the currents from seven multipliers are added together and converted to voltage by the termination resistor, the IR drop across the resistor would be very large, forcing the multiplier transistors into the triode region. For example, in this design, the current of each multiplier is 4 mA and the load resistance (constrained by the transmission line) is 45 Ω. Therefore, the multiplier common mode voltage would be 370 mV, which is too low for keeping the multiplier differential pair in the saturation region. Second, a large amount of current, 14 mA, is conducted in the passive delay line. In order to prevent electromigration issues, the metal interconnect needs to be sufficiently wide causing extra parasitic capacitance and lowing overall bandwidth.

The ISI compensation of most channels requires smaller post- and pre-cursor taps than the main tap. By scaling down their gain, the required current in these taps can be reduced. The resulting current reduction not only lowers the chip power consumption but also helps with IR drop and electro-migration issues at the summer node. To this end, the gain of taps 2, 3, 5 and 6 is reduced by 50%; the gain of taps 1 and 7 is reduced by 75%. The scaling lowers the total multiplier/summer current consumption from 28 mA to 14 mA. If a link requires higher pre- or post-cursor taps than the ones provided, the main tap weight can be decreased to increase the relative weight of other taps.

Fig. 14 shows the 50% scaled multiplier where only the transistor multiplier factor M is reduced by a factor of 2; the transistor sizes remain the same. This approach leads to better matching between the scaled and non-scaled versions. Because the delay of all the taps needs to be equal, the gain scaling should not impact the tap delay; the input and output capacitances of the multiplier need to remain constant. Dummy transistors are added to scaled multipliers to maintain these

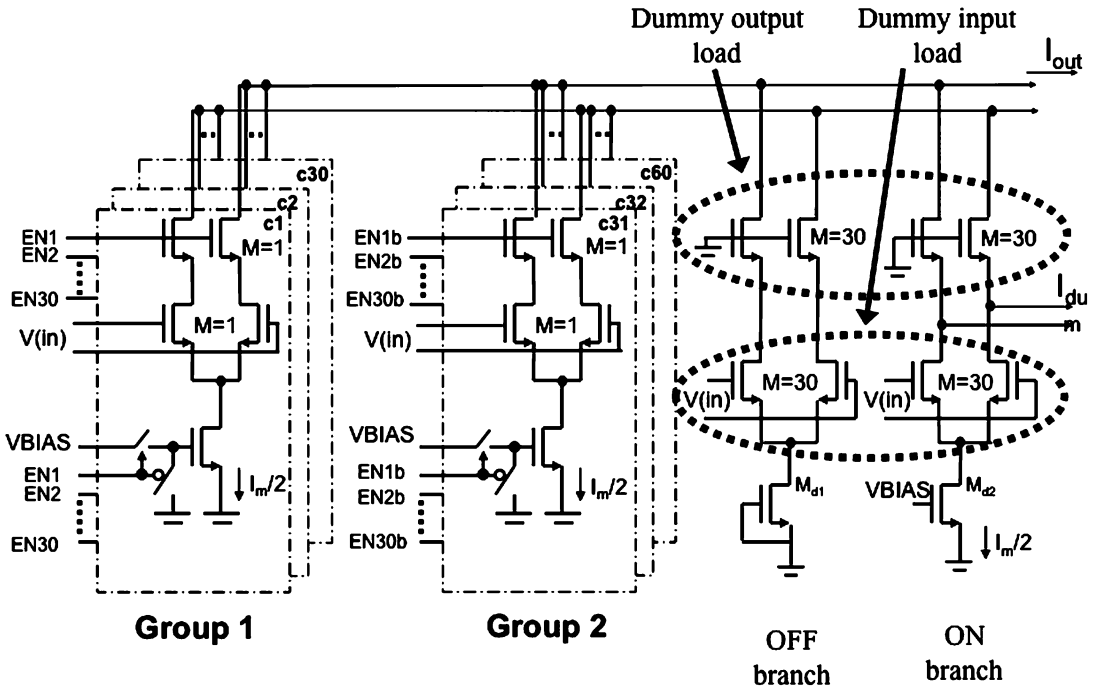

Fig. 14. 50% scaled multiplier.

capacitances. But since the transistor gate capacitance depends on whether it is on or off, the scaled multiplier contains two dummy branches: one for transistors that are turned on and the other transistors that are turned off. As mentioned previously, for all gain settings the full-scale multiplier always contains 30 on and 30 off unit cells each with $M = 2$. As a result, a total of 60 transistors are on and 60 are off. In the unit cell of the half-scaled version, $M = 1$, resulting in total of 30 on transistors and 30 off transistors. Hence, 30 dummy on transistors and 30 dummy off transistors are added as shown in Fig. 14. The tail transistor M_{d1} is shut off by grounding its gate; the gate of M_{d2} is tied to the multiplier bias line VBIAS. This design guarantees the input capacitance matching of the scaled and non-scaled version.

To match the output capacitances, dummy cascode transistors are added to both branches. Similar reasoning as above indicates that M should be set to 30 for the dummy cascode transistors of the on and off branches. It should be noted that the current in the ON dummy branch cannot be added to the multiplier output and hence a different current output path has been generated. The dummy output currents $I_{\rm dum}$ of all the scaled multipliers are added together then sourced through a resistor connected to $V_{\rm DD}$. Using a similar approach a 75% scaled multiplier is designed.

In order to match the delay of the 7th tap with the other taps, a dummy delay element has been added to its output. Although its circuit topology is similar to the active delay cell, all the inductors have been removed to save area and its current has been reduced by a factor of 20.

Finally, an additional active delay element is added before the first tap. This additional stage ensures that the input common-mode voltage and rise/fall time for all the taps are similar, hence reducing the tap delay mismatch even further.

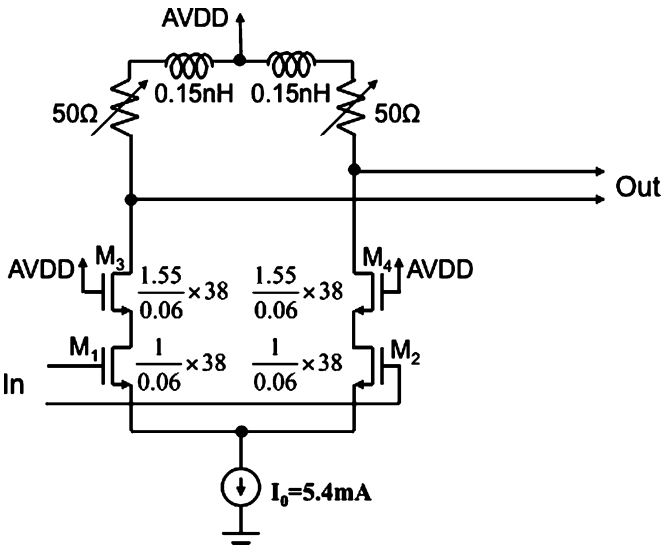

Fig. 15. Output driver schematic.

E. Input Termination and Output Drivers

The design of the 40 Gb/s input and output paths depends heavily on the package. This chip was packaged in a flip-chip ball grid array (BGA) where on-die bumps provide the electrical connection between the die and the package. The required spacing between the bumps is dictated by the package, which leads to 150-μm-long interconnect (approx. 150 pH of inductance) between the bump and the chip 40 Gb/s I/O. On-chip transmission lines are used for these long interconnects; matching microstrip lines with large bandwidths are the best choice. Assuming the characteristic impedance Z_0 can be approximated by $\sqrt{L/C}$, then $C = L/Z_0^2$. Thus, the required capacitance C is 57 fF. Using these values, the microstrip lines

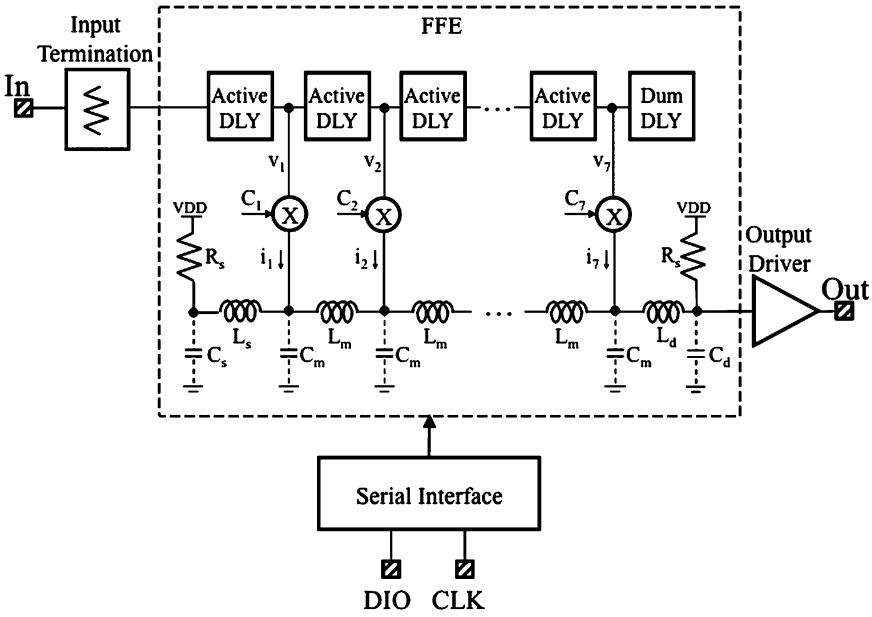

Fig. 16. The chip block diagram.

are implemented by routing the signal through a 6 μm-wide metal-7 line over a 26 μm-wide metal-1 ground plane. The HFSS simulation result indicates that the bandwidth of the designed microstrip lines is greater than 50 GHz.

ESD diodes were also added to all the 40 Gb/s bumps to protect the chip against possible electrostatic discharge during testing or handling. The size of the diodes has been minimized so that their capacitive contribution is less than 10 fF.

In the output path, the driver in Fig. 15 is used. The cascode transistors have been implemented similar to the active delay elements. Calibrated 50 Ω resistors provide the output termination. The shunt-peaking inductors have been optimized for maximum bandwidth.

Fig. 16 shows the chip top-level block diagram where all the differential signals are represented by single-ended connections for simplicity. The incoming 40 Gb/s data is applied to the 100 Ω differential termination block. The FFE core equalizes the received ISI and its output is transmitted out of the chip through the 50 Ω output driver. Various adaptation algorithms, such as Least Mean Square, Zero Forcing, and dithering can be used to adapt the FFE coefficients. In this chip, the algorithm is not implemented on-chip and the FFE coefficients are manually programmed through the chip serial interface. The FFE core occupies 0.75 mm^2 in a 65 nm CMOS process and consumes 65 mW from a 1 V supply, making it the lowest power consuming FFE published so far (Table I). The power consumption of the entire chip, including the FFE, the input termination, serial interface and 50 Ω output driver, is 80 mW. The die photo is shown in Fig. 17.

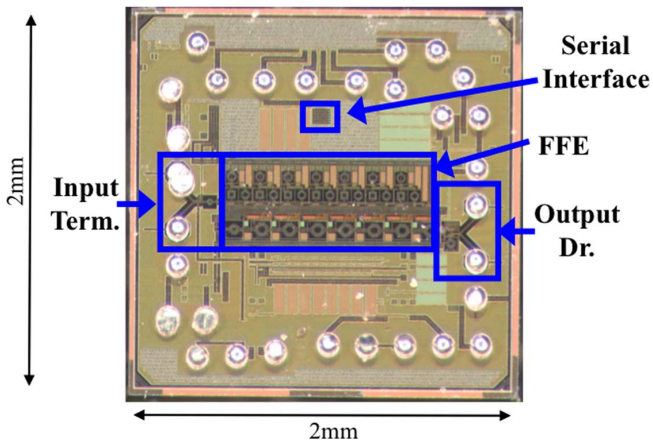

Fig. 17. Die photo.

III. Circuit Measurements

The performance of the package and on-chip termination is characterized by the output return loss parameter S22. As mentioned in the previous section, the high-speed input and output paths include the package, ESD structure, on-chip transmission line, and 100 Ω differential termination. In an effort to separate the contribution of the package and the die to the return loss results, both the bare die and the packaged parts were measured. In addition, to build further confidence in the results the part-to-part variation has been measured on three different bare die and three different packaged parts. Fig. 18 shows the measured results for both bare die and the packaged parts. Although the package seems to be the limiting factor for the chip return loss performance, the S22 is smaller than -10 dB up to 30 GHz. Furthermore, the obtained results from different die and packaged devices are similar, suggesting robust performance.

The measured frequency response of the chip and a single delay tap are shown in Fig. 19. By setting all the FFE coefficients to zero except for the main tap (C_4), the transfer function of the chip with no equalization is obtained. As shown in Fig. 19, the measured 3dB bandwidth is larger than 20 GHz. Next, C_4 is set to zero, C_5 is maximized, and the chip transfer function is remeasured. The difference between the two transfer functions corresponds to the transfer function of one delay element (5th

TABLE I
PERFORMANCE COMPARISON

	This work	[9]	[10]	[11]	[15]
# taps	7	3	4	3	7
Tap delay (ps)	12.5	25	12.5	--	6.75
Total delay (ps)	87.5	75	50	--	47.3
Technology	65 nm CMOS	0.18 µm CMOS	InP	InP/InGaAs	SiGe
Supply (V)	1	1.8	-5.2	-4.3	5
FFE Power (mW)	65[1]	70	1500	820	750
Package	Flip-chip BGA	None	Module	--	None

[1] Total chip power is 80mW.

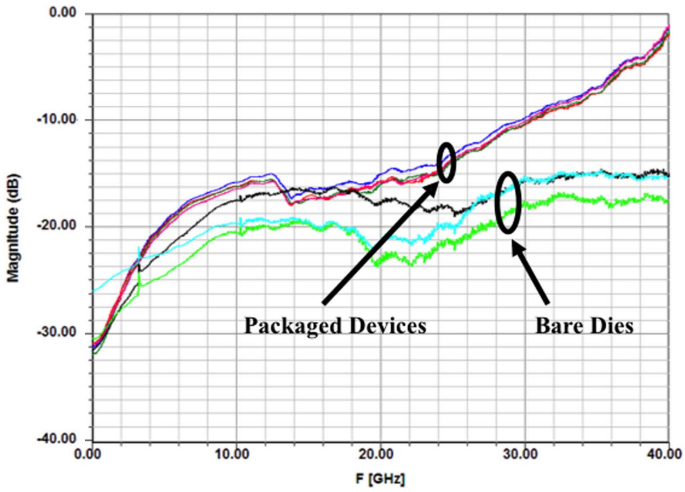

Fig. 18. The measured returned loss data for the bare die and packaged devices.

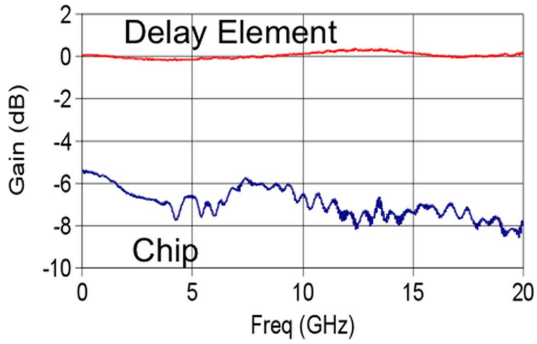

Fig. 19. Measured frequency domain results of the delay cell and the full chip.

tap). The measurements shown in Fig. 19 verify that the delay element exhibits a flat response with less than 1 dB of variation up to 20 GHz.

Fig. 20 shows the measured time-domain response of each individual tap with a 5 GHz sine wave input. The measured time delay is about 15.3 ps with less than 0.3 ps of variation across all seven taps suggesting that the input and capacitance of the

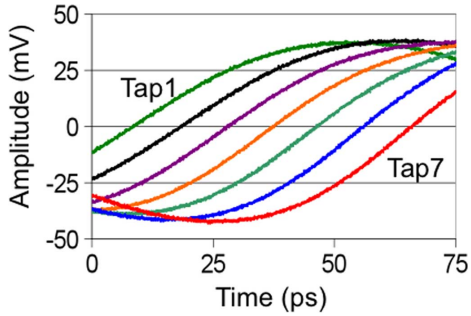

Fig. 20. Measured time-domain results of each tap. The gain and delay are well matched for all 7 taps for this design.

scaled and non-scaled multipliers are well matched. In addition, the output amplitude of each tap is relatively equal which shows that the gain of the delay element is very close to unity. In contrast, the delay element performance of a previously published work [9] exhibits larger gain and delay variation.

The equalization capability of the chip was measured in the time domain; the input and output eye diagrams are shown in Fig. 21. A 40 Gb/s PRBS31 data is passed through 4 inches of FR4 traces with 9 dB attenuation at 20 GHz, creating the closed eye. By optimizing the equalizer coefficients, the open eye diagram is achieved with more than 50% and 70% vertical and horizontal openings, respectively, corresponding to approximately 7.5 ps p-p of total jitter.

Table I gives a comparison of the performance of this chip with previously published 40 Gb/s FFE equalizers.

IV. CONCLUSION

We have presented the design and measurement of a 7-tap feed-forward equalizer. The chip was manufactured using the TSMC standard 65 nm CMOS process and includes input termination and a 50 Ω output driver in addition to the equalizer while consuming only 80 mW of power. By adding ESD protection to 40 Gb/s IOs and utilizing inexpensive 6 mm × 6 mm plastic flip-chip BGA packaging, the chip is rendered closer to a commercial product.

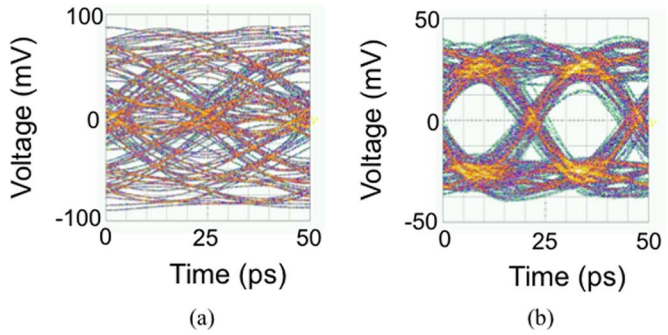

Fig. 21. Measured 40 Gb/s eye diagrams: (a) before equalization; (b) after equalization.

The chip's performance was achieved through various architectural and circuit design techniques. The FFE tap delay was broken into a dominant active and a small passive portion where the effective delay is the difference between those of the two individual blocks. This approach eases the input/output signal integrity issue while reducing equalizer sensitivity to inductor modeling inaccuracy. In the active delay element design, various broad-banding techniques, such as shunt- and series-peaking, along with resistor and temperature compensation were used. A switchable multiplier structure was shown to improve the gain step size uniformity, reduce the distortion and increase the maximum gain without increasing the power consumption. The proposed tap scaling was shown to reduce the overall power consumption while easing the issues related to 65 nm low power supply level. Finally, due to its design in a CMOS process, the FFE can be integrated with a CDR and Demux, eliminating the power hungry high-speed chip-to-chip connection. These integrations can help EDC become more practical and transform the highly dispersed 40 Gb/s optical link to a mainstream communication media.

REFERENCES

[1] H. Wu, J. A. Tierno, P. Pepeljugoski, J. Schaub, S. Gowda, J. A. Kash, and A. Hajimiri, "Integrated transversal equalizers in high-speed fiber-optic systems," *IEEE J. Solid-State Circuits*, vol. 38, no. 12, pp. 2131–2137, Dec. 2003.

[2] H. Jiang, R. Saunders, and S. Colaco, "SiGe equalizer IC for PMD mitigation and signal optimization of 40 Gbits/s transmission," presented at the Optical Fiber Communication Conf., OFC/NFOEC, Apr. 2005, OWO2.

[3] G. Zhang, P. Chaudhari, and M. M. Green, "A BiCMOS 10 Gb/s adaptive cable equalizer," in *2004 IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2004, pp. 748–749.

[4] S. Gondi, J. Lee, D. Takeuchi, and B. Razavi, "A 10 Gb/s CMOS adaptive equalizer for backplane applications," in *2005 IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2005, pp. 328–329.

[5] R. Gitlin and S. Weinstein, "Fractionally-spaced equalization: An improved digital transversal equalizer," *Bell Syst. Tech. J.*, vol. 60, no. 2, pp. 275–296, Feb. 1981.

[6] P. Watts, V. Mikhailov, S. Savory, P. Bayvel, M. Glick, M. Lobel, B. Christensen, P. Kirkpatrick, S. Shang, and R. Killey, "Performance of single-mode fiber links using electronic feed-forward and decision feedback equalizer," *IEEE Photon. Technol. Lett.*, vol. 17, no. 10, pp. 2206–2208, Oct. 2005.

[7] A. Momtaz and M. Green, "An 80 mW 40 Gb/s 7-tap T/2-spaced FFE in 65 nm," in *2009 IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2009, pp. 364–365.

[8] C. Pelard, E. Gebara, A. J. Kim, M. G. Vrazel, F. Bien, Y. Hur, M. Moonkyun, S. Chandramouli, C. Chun, S. Bajekal, S. E. Ralph, B. Schmukler, V. M. Hietala, and J. Laskar, "Realization of multigigabit channel equalization and crosstalk cancellation integrated circuits," *IEEE J. Solid-State Circuits*, vol. 39, no. 10, pp. 1659–1670, Oct. 2004.

[9] J. Sewter and A. C. Carusone, "A 3-tap FIR filter with cascaded distributed tap amplifiers for equalization up to 40 Gb/s in 0.18 μm CMOS," *IEEE J. Solid-State Circuits*, vol. 41, no. 8, pp. 1919–1929, Aug. 2006.

[10] S. Wada, R. Ohhira, T. Ito, J. Yamazaki, Y. Amamiya, H. Takeshita, A. Noda, and K. Fukuchi, "Compensation for PMD-induced time-variant waveform distortions in 43-Gbit/s NRZ transmission by ultra-wideband electrical equalizer module," presented at the Optical Fiber Communication Conf., OFC/NFOEC, Apr. 2005, OWE2.

[11] M. Nakamura *et al.*, "Electrical PMD equalizer ICs for a 40 Gb/s transmission," presented at the Optical Fiber Communication Conf., OFC/NFOEC, Apr. 2004, TuG4.

[12] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*, 2nd ed. New York: Cambridge Univ. Press, 2004, pp. 325–327.

[13] D. Chung, "Resistor compensation apparatus," U.S. Patent 7,042,271, May 9, 2006.

[14] S. Reynolds, P. Pepeljugoski, J. Schaub, J. Tierno, and D. Beisser, "A 7-tap transverse analog-FIR filter in 0.13 μm CMOS for equalization of 10 Gb/s fiber-optic data systems," in *2005 IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, 2005, pp. 330–331.

[15] A. Hanzneci and S. Voinigescu, "A 49-Gb/s, 7-tap transversal filter in 0.18 μm SiGe BiCMOS for backplane equalization," in *Proc. 2004 IEEE Compound Semiconductor Integrated Circuit Symp. (CSICS)*, Oct. 2004, pp. 101–104.

Afshin Momtaz (S'89–M'90) received the M.S.E.E. degree from the University of California, Los Angeles (UCLA), and the Ph.D. degree from the University of California, Irvine.

From 1992 to 1998, he was with Western Digital and Adaptec corporations where he designed disk-drive read channel and fiber channel transceiver chips. In 1998, he joined NewPort Communications, where he was involved with design of various OC48 and OC192/10GE CMOS transceivers. Since the acquisition of NewPort Communications by Broadcom in 2000, he has been also focusing on 10 Gbps equalizer designs for optical, copper and backplane applications. He is currently a Senior Design Manager of the Analog and Mixed Signal group and a Broadcom Distinguished Engineer. He has authored or coauthored more than 15 journal/conference papers, and holds more than 55 U.S. patents issued in the area of multi-GHz mixed-signal circuits and systems.

Dr. Momtaz was the co-chair of the Wireline Communication subcommittee of the IEEE Custom Integrated Circuit Conference 2009.

Michael M. Green (S'89–M'91) received the B.S. degree in electrical engineering from the University of California, Berkeley, in 1984 and the M.S. and Ph.D. degrees in electrical engineering from the University of California, Los Angeles (UCLA), in 1988 and 1991, respectively.

He has been with the Department of Electrical Engineering and Computer Science at the University of California, Irvine, since 1997, where he is currently Professor and Chair. From 1999 to 2001 he was an IC designer with the Optical Transport Group at Broadcom (formerly NewPort Communications). His current research interests include the design of analog and mixed-signal integrated circuits for use in high-speed broadband communication networks and nonlinear circuit theory. He has published over 50 papers in technical journals and holds three patents.

Dr. Green was the recipient of the Outstanding Master's Degree Candidate Award in 1989 and the Outstanding Ph.D. Degree Candidate Award in 1991, both from the UCLA School of Engineering and Applied Science. He is also the recipient of the Sigma Xi Prize for Outstanding Graduate Science Student at UCLA in 1991, the 1994 Guillemin–Cauer Award of the IEEE Circuits and Systems Society, the 1994 W. R. G. Baker Award of the IEEE, a 1994 National Young Investigator Award from the National Science Foundation and the Award for New Technical Concepts in Electrical Engineering from IEEE Region 1. He has served as an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS, IEEE TRANSACTIONS ON VLSI, and IEEE TRANSACTIONS ON EDUCATION.