8.1 Circuit Techniques for a 40Gb/s Transmitter in 0.13μm CMOS

Jaeha Kim[1], Jeong-Kyoum Kim[1], Bong-Joon Lee[1], Moon-Sang Hwang[1], Hyung-Rok Lee[1], Sang-Hyun Lee[2], Namhoon Kim[2], Deog-Kyoon Jeong[2], Wonchan Kim[1]

[1]Seoul National University, Seoul, Korea
[2]Silicon Image, Sunnyvale, CA

A 40Gb/s transmitter, implemented in a 0.13μm CMOS process with f_T of 70GHz, is presented. Since such a high speed is well beyond the reach of conventional CMOS designs, circuit techniques that use inductive peaking, negative feedback, and pulsed latches are presented to achieve the bandwidth and timing required for transmitting a 40Gb/s NRZ bit stream.

The proposed 40Gb/s transmitter has a tree structure that performs 2:1 multiplexing at each level of hierarchy, as shown in Fig. 8.1.1. A 20GHz VCO provides a 50%-duty clock for the final MUX that selects which of the two inputs drives the output. The clock is divided down by a chain of clock dividers to generate clocks for the MUXs at the subsequent levels. Retiming latches realign the signals as they enter each different clock domain. A PFD and a loop filter (LF) close a loop to lock the divided-by-32 VCO clock to a 625MHz reference. CML circuits are used wherever the frequency is above 2GHz.

The critical components in this transmitter are: the buffer amplifiers for the 20GHz clock and 40Gb/s random bit stream, the 2:1 MUX for serializing the 40Gb/s bit stream, and the retimers and dividers operating at 20GHz.

First, the CML buffers use shunt and double-series peaking for bandwidth enhancement [1]. As shown in Fig. 8.1.2, shunt and double-series peaking consists of 3 inductors forming a T-shape among the driver transistor, the load resistor, and the output node. These inductors delay the current flows so that the drain junction capacitance, the parasitic capacitances of the inductors and resistor, and the load capacitance are charged serially in time and the driver never sees them at once. This improves the signal rise time at the cost of increased latency (Fig. 8.1.4a).

The T-bridged inductors are equivalent to a transformer. Therefore, they are implemented as a single planar spiral transformer (Fig. 8.1.2). Hence, area and parasitics of 3 separate inductors are saved. Notice that the inner and outer spirals have opposite-turn directions to realize a positive coupling coefficient of k. A negative k would make a transformer for a bridged T-coil, which is found less effective than shunt-series peaking as in [2].

Finding the layout geometry that yields the best performance is challenging, as the choice of L_1, L_2, and k is constrained by layout feasibility and the resulting parasitics alter the optimal solution. We numerically optimized the transformer geometry via simulated annealing, by using HSPICE to refine choices for L_1, L_2, and k and ASITIC to find the best geometry and extract the parasitics. The transistor nonlinearity and large internal swing of 500mV make small-signal parameters such as bandwidth and group delay poor goals for optimization. Better results are obtained by minimizing the rise time and overshoot of the step response. Surprisingly, the numerically tuned inductances are quite different from those predicted by theory [1, 2] and the improvements sometimes exceeded the theoretical bounds. In CML buffers, shunt-peaking extends the –3dB bandwidth from 13GHz to 24GHz (1.83×) and shunt and double-series peaking extends it to 42GHz (3.23×), while keeping the frequency response flat.

The MUX in Figure 8.1.3 uses negative feedback, since the inductive peaking alone cannot achieve enough bandwidth for 40Gb/s due to the large junction capacitance where the two differential pairs merge. This form of feedback can improve the gain-bandwidth product by a factor of f_T/BW and is more amenable to low-voltage operation than the Cherry-Hooper topology [3].

An alternative way of explaining the increased bandwidth is to think of the feedback amplifier as a de-emphasizing equalizer. The amplifier injects an opposite-signed, scaled-down, and delayed version of the main-stage current to the bandwidth-limiting node of the MUX. When the node starts switching, this feedback current helps the transition, enhancing the high-frequency gain. Once the node settles, the feedback current changes its polarity and fights against the main current, reducing the low-frequency gain. As a result, the frequency response becomes flat for a broader range but has a lower gain (Fig. 8.1.4b). Simulation shows that negative feedback extends the –3dB bandwidth from 16.5 to 27GHz, compared to when only shunt peaking is used.

The CML latches in the retimers and clock dividers also use inductive peaking and negative feedback to extend the bandwidth. However, the bandwidth enhancement comes at the cost of increased latency. The master-slave FF, made of these latches, has a latency of 48ps. Thus, the retimers and clock dividers based on FFs will have difficulty operating beyond 20GHz.

The proposed 20GHz retimer and clock divider use pulsed-latches to mitigate this long latency. Recall that in a master-slave FF, the role of the master latch is to keep the input from racing through while the slave latch is transparent. Since the latency of the slave already precludes race-through at 20GHz, the master can be safely removed, thus saving time. The resulting retimer and clock divider are shown in Fig. 8.1.1. These circuits have a lower bound on operating frequency, since the fixed latency cannot prevent race-through at low frequencies. However, the operating range of the proposed clock divider is 13 to 26GHz, much wider than the VCO frequency range. The 20GHz pulsed-latch-based retimer does not further limit the operating range.

A 20GHz clock calls for a resonant oscillator, but the difficulty lies in accurately predicting its frequency range. With tank capacitance of 1.2pF, the inductance that yields 20GHz is only 0.053nH and the required precision is hard to meet with spiral inductor models. Therefore, this inductance is realized with coupled microstrips, whose geometry and VCO circuits are shown in Fig. 8.1.5. Once the 2D geometry is characterized, the effective inductance can be fine-tuned by adjusting the length. The length of the microstrip and the sizes of the transistors and varactors are optimized via the guidelines of [4]. The measured VCO tuning range is 17.6 to 19.4GHz, close to the simulated range of 18 to 20GHz. The simulated Q at 20GHz is 23, limited by the series loss of the varactors rather than that of the microstrips. The 20GHz clock jitter measured with a 1010 fixed data pattern at the TX output is $4.9ps_{pp}$ and $0.65ps_{rms}$.

The prototype TX chip is fabricated in a 0.13μm CMOS MS/RF 1P6M process that provides 3.3μm-thick top metal and varactors. The chip occupies 2.5×3.6mm² and contains 410 inductors and transformers. The CML circuits dissipate 2.7W and the other CMOS circuits 107mW while operating at 1.5V and 1.3V, respectively. Figure 8.1.6 shows the differential eye diagrams of the TX, probed directly on the bond pads, when the reference clock is 600MHz. The transmit jitter of the 2^{31}-1 PRBS 38.4Gb/s eye is as low as $8.11ps_{pp}$ and $1.53ps_{rms}$ and the 10%-to-90% rise-time is 14ps, demonstrating the feasibility of 40Gb/s in CMOS.

References

[1] T. H. Lee, "The Design of CMOS Radio-Frequency Integrated Circuits," Cambridge University Press, New York, NY, 1998.
[2] S. Galal, B. Razavi, "40Gb/s Amplifier and ESD Protection Circuit in 0.18μm CMOS Technology," *ISSCC Dig. Tech. Papers*, pp. 480-481, Feb., 2004.
[3] S. Galal, B. Razavi, "10Gb/s Limiting Amplifier and Laser/Modulator Driver in 0.18μm CMOS Technology," *ISSCC Dig. Tech. Papers*, pp. 188-189, Feb., 2003.
[4] D. Ham, A. Hajimiri, "Concepts and Methods in Optimization of Integrated LC VCOs," *IEEE J. Solid-State Circuits*, vol. 36, no. 6, pp. 896-909, Jun., 2001.

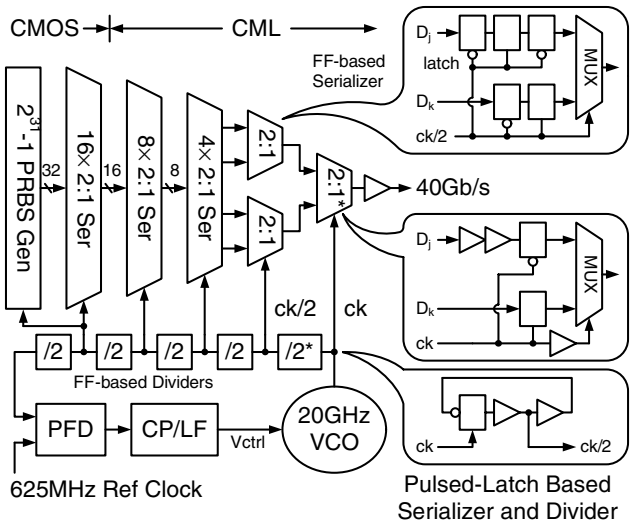


Figure 8.1.1: 40Gb/s transmitter architecture including pulsed-latch-based datapath.

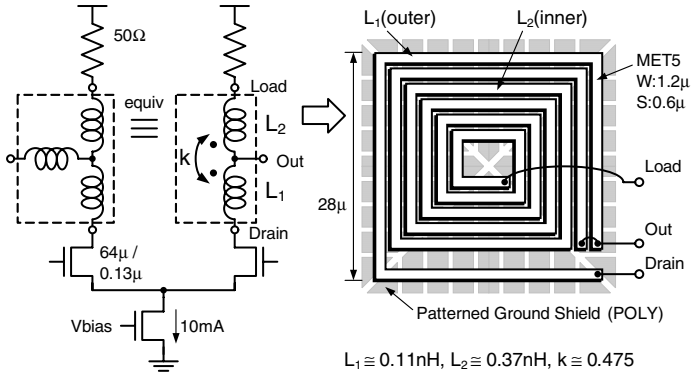


Figure 8.1.2: Bandwidth enhancement using shunt and double-series peaking implemented as a planar spiral transformer.

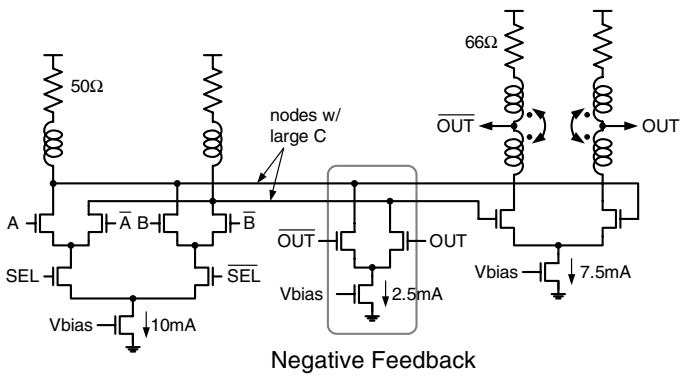
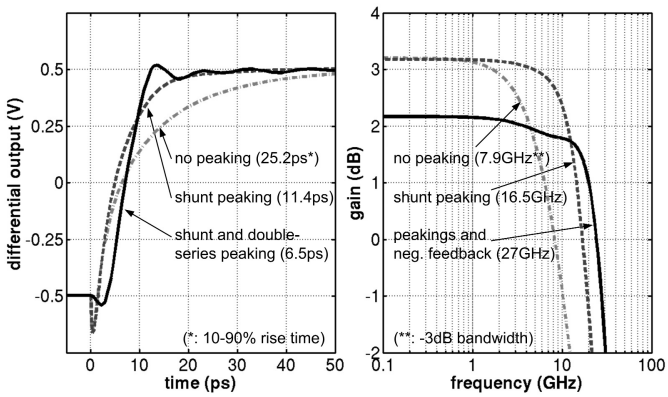


Figure 8.1.3: 2:1 CML MUX with inductive peaking and negative feedback.



(a) Step response of a buffer with inductive peaking

(b) Frequency response of 2:1 multiplexer (SEL-to-OUT)

Figure 8.1.4: Simulated responses of the proposed CML buffer and 2:1 MUX.

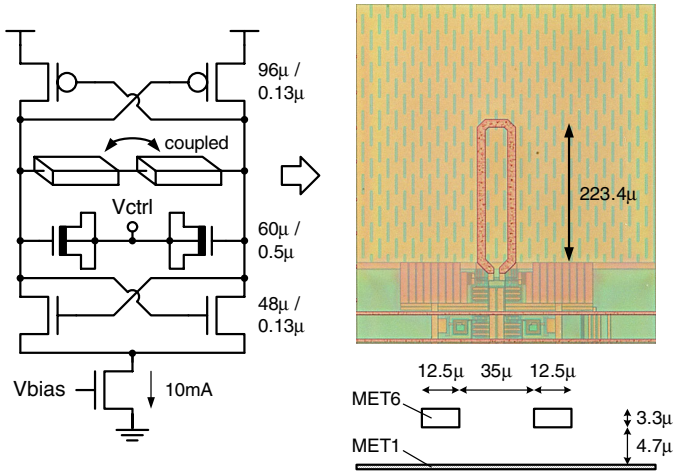
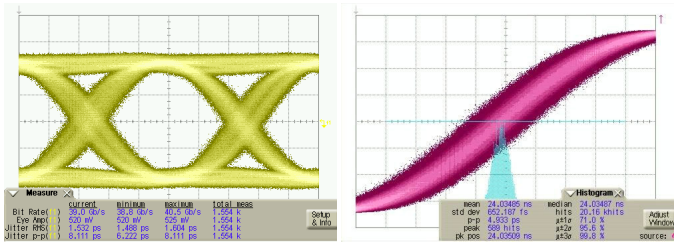


Figure 8.1.5: 20GHz VCO with a coupled microstrip resonant tank.



H: 5.0ps/div, V: 145mV/div
Jitter: 8.11ps p-p, 1.53ps rms
Amp: 520mV p-p differential

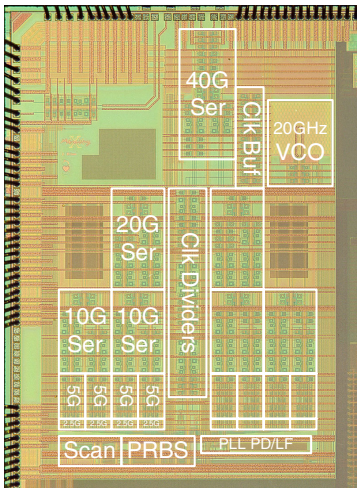
(a) Eye diagram of 38.4Gb/s transmit data

H: 2.0ps/div, V: 50mV/div
Jitter: 4.93ps p-p, 652 fs rms

(b) Jitter histogram of 19.2GHz transmit clock

Figure 8.1.6: Measured eye diagram of 2^{31}-1 PRBS TX data at 38.4Gb/s.

Continued on Page 589



Process	0.13µm CMOS MS/RF
Area	2.5 × 3.6 mm^2
Supply	1.3V (CMOS), 1.5V (CML)
Power	107mW (CMOS), 2.7W (CML)
Bitrate	35.2 ~ 38.7 Gbps
VCO lock range	17.6 ~ 19.4 GHz
Ref. clock	600MHz (@ 38.4Gbps)
Tx jitter	8.11ps p-p, 1.53ps rms
Tx swing	520mV p-p differential
# Inductors	410

Figure 8.1.7: Chip micrograph and performance summary.

