

40-Gb/s Amplifier and ESD Protection Circuit in 0.18- μm CMOS Technology

Sherif Galal and Behzad Razavi, *Fellow, IEEE*

Abstract—A triple-resonance LC network increases the bandwidth of cascaded differential pairs by a factor of $2\sqrt{3}$, yielding a 40-Gb/s CMOS amplifier with a gain of 15 dB and a power dissipation of 190 mW from a 2.2-V supply. An ESD protection circuit employs negative capacitance along with T-coils and pn junctions to operate at 40 Gb/s while tolerating 700–800 V.

Index Terms—Broadband amplifiers, distributed amplifiers, ESD, inductive peaking, resonant circuits, T-coils.

I. INTRODUCTION

THE continuous growth of multimedia communications has been driving data transmission systems to higher speeds. With the evolution of wavelength division multiplexing (WDM), data rates of 40 Gb/s became possible. This requires electronic systems operating at these high data rates to interface with the optical medium. Recently, 40-Gb/s circuits realized in SiGe have been reported [1]–[3].

Fig. 1 shows a typical 40-Gb/s transceiver. In the transmit path, the laser/modulator driver uses very large transistors and itself requires a heavy input drive. Thus, a predriver must be interposed between the retiming flipflop and the driver. In the receive path, the transimpedance amplifier (TIA) bandwidth requirements translate to a relatively low gain, necessitating a postamplifier to overcome the noise of the subsequent equalizer.

For 40-Gb/s applications, the required bandwidth is around 26 GHz¹ and the required gain on the order of 10–15 dB to avoid degrading the receiver sensitivity due to the noise of the equalizer. This paper describes a broadband technique in the context of an amplifier design that can precede or follow an equalizer in the receiver or act as a predriver in the transmitter front end. An ESD circuit is also presented that can be used for input or output nodes of 40-Gb/s circuits.

The next section describes the issues in the design of distributed amplifiers (DAs) in CMOS technology. Section III presents the proposed amplifier, Section IV presents the ESD protection circuit, and Section V summarizes the experimental results.

II. DISTRIBUTED AMPLIFICATION TECHNIQUES

While distributed circuits have been considered as an attractive candidate for high-speed amplification, several issues

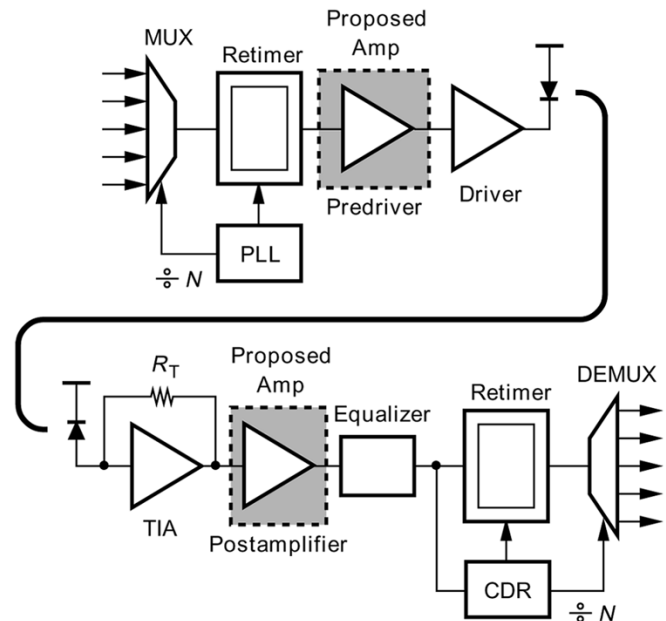


Fig. 1. Typical 40-Gb/s transceiver.

make their realization in CMOS technology difficult (Fig. 2). First, since the bias currents of all of the stages flow through the same loads, the circuit suffers from a severe tradeoff between voltage gain and voltage headroom [Fig. 2(a)], especially if MOSFETs are biased at a high current density to maximize their f_T . In other words, the gain–headroom tradeoff here is much more serious than in cascaded stages. Second, the loss of the transmission lines in CMOS technology limits the length of the line, the number of sections, and ultimately the gain that can be achieved.

Third, the finite output resistance of short-channel transistors yields additional loss in the output transmission line [Fig. 2(b)]. Specifically, if m transistors are used per unit length of the transmission lines, then the loss factor is given by [4]

$$\alpha = \frac{1}{2} \left(\frac{R_s}{Z_o} + \frac{m}{r_o} Z_o \right) \quad (1)$$

where R_s denotes the series resistance per unit length, Z_o denotes the characteristic impedance, and r_o denotes the output resistance of each transistor. To minimize the loss, the parallel combination of all output resistances must be much higher than the line characteristic impedance. For example, in a five-stage design employing sufficiently high bias currents, to achieve a gain of 15 dB with a characteristic impedance of 50 Ω , each transistor suffers from an output resistance of only 1 k Ω , and the second term in the above equation becomes

Manuscript received April 16, 2004; revised July 16, 2004.

The authors are with the Electrical Engineering Department, University of California, Los Angeles, CA 90095-1594 USA (e-mail: razavi@icsl.ucla.edu).
Digital Object Identifier 10.1109/JSSC.2004.835639

¹The bandwidth of the amplifier must be larger to accommodate the limited bandwidth of the TIA and the equalizer.

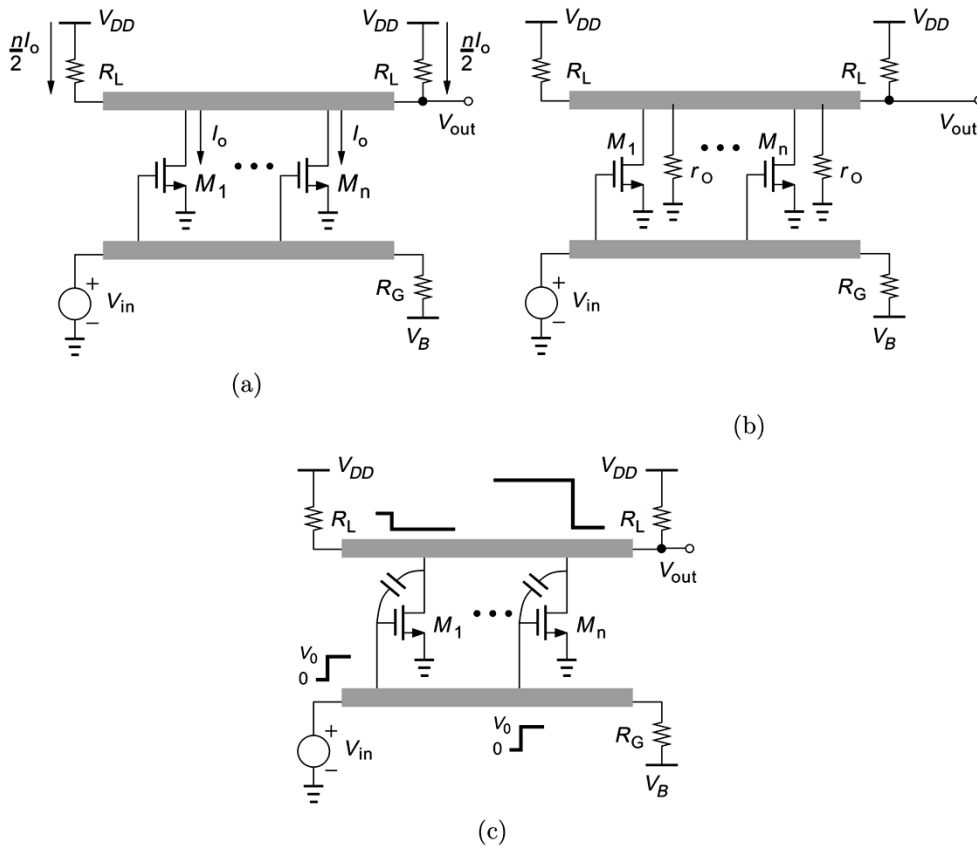


Fig. 2. Performance limitations of DAs: (a) voltage headroom, (b) output resistance of transistors, and (c) nonuniform Miller effect.

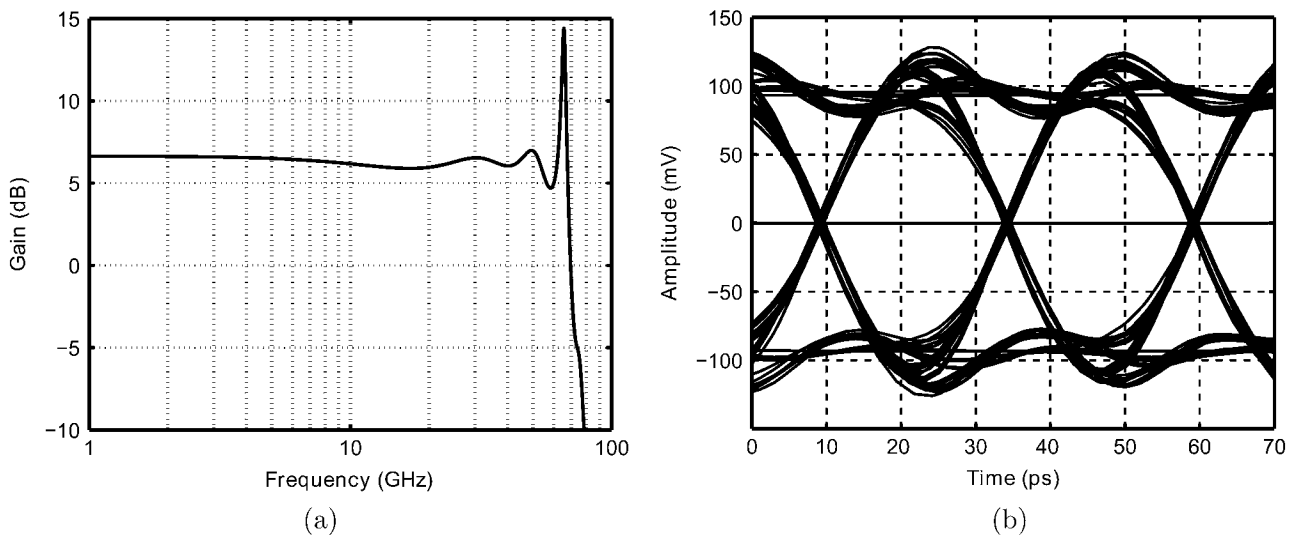


Fig. 3. Simulated (a) frequency response and (b) output eye of a five-stage DA.

comparable with the first. This issue continues to worsen in deep-submicron generations.

Fourth, the gate-drain overlap capacitance of the transistors experiences *nonuniform* Miller multiplication because the apparent gate-to-drain gain increases, as shown in Fig. 2(c). As a result, the effective capacitance per unit length of the input transmission line increases from left to right, lowering both the characteristic impedance and the wave velocity and potentially introducing ISI in random data. Since the n th transistor sees a

gain n times that experienced by the first transistor, and since $C_{GD} \approx 0.25C_{GS}$, we note that a gain of six translates to a 35% variation in $Z_o = \sqrt{L_o/C_o}$ from one end to the other.

As an example, a five-stage 40-Gb/s differential DA has been designed in 0.18- μ m CMOS technology. Fig. 3 plots the simulated frequency response² and the output eye of the amplifier, exhibiting a maximum gain of only 6.6 dB.

²Note that even ideal DAs exhibit significant peaking near the cut-off frequency of the lines [5].

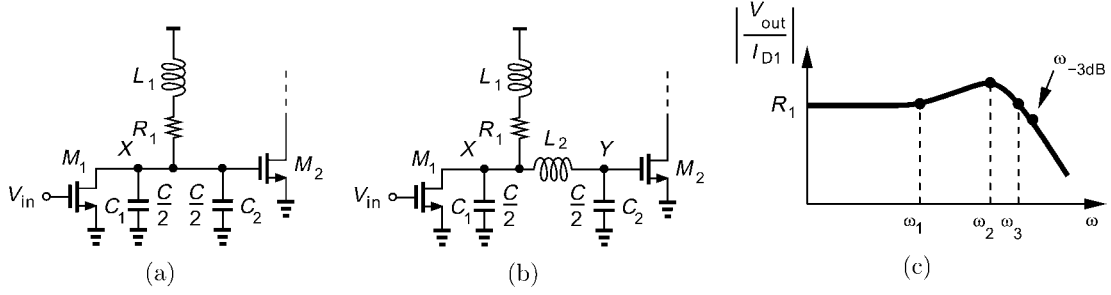


Fig. 4. (a) Inductively peaked stage, (b) TRA, and (c) frequency response of the TRA.

The above shortcomings arise because of the additive nature of the gain in DAs. On the other hand, multiplicative gain and hence cascaded stages do not face these difficulties but require a large bandwidth per stage. The next section introduces a technique that raises the bandwidth of cascaded differential pairs by a factor of $2\sqrt{3} \approx 3.5$, well above the factors corresponding to inductive or T-coil peaking.

III. 40-Gb/s AMPLIFIER

A. Triple-Resonance Architecture

To arrive at the concept of the triple-resonance amplifier (TRA), first consider the inductively peaked cascade of two stages shown in Fig. 4(a), where it is assumed that M_1 and M_2 contribute approximately equal capacitances ($C/2$) to node X . As the frequency approaches

$$\omega_1 = \frac{1}{\sqrt{L_1 C}} \quad (2)$$

the impedance of L_1 rises, allowing a greater fraction of I_{D1} to flow through $C_1 + C_2$ and hence extend the bandwidth.

To increase the bandwidth, we insert an inductor L_2 in series with C_2 [Fig. 4(b)] such that L_2 and C_2 resonate at ω_1 , thereby acting as a short and absorbing all of I_{D1} . Now, I_{D1} flows through C_2 rather than $C_1 + C_2$, leading to a more gradual roll-off of gain. For L_2 and C_2 to resonate at ω_1 , we have

$$L_2 = 2L_1. \quad (3)$$

(Since, in practice, C_1 and C_2 are not exactly equal, the ratio of L_1 and L_2 can be slightly adjusted to compensate for this difference.) To minimize peaking, the output voltage at this frequency $I_{in}/(C_2\omega_1)$ must be equal to that at low frequencies $I_{in}R_1$, yielding

$$R_1 = 2\sqrt{\frac{L_1}{C}}. \quad (4)$$

The amplifier exhibits the frequency response shown in Fig. 4(c), revealing three distinct resonance frequencies. For this reason, we call this topology a ‘‘triple-resonance amplifier’’ (TRA).³ To understand the operation of the TRA and derive the required relationships among different components, the behavior of the amplifier will be studied around each of these frequencies.

³A similar topology is described in [6] but requiring $L_2 = 4.3L_1$ and $C_2 = 2C_1$. It is also unclear if [6] exploits the three resonance frequencies to broaden the bandwidth.

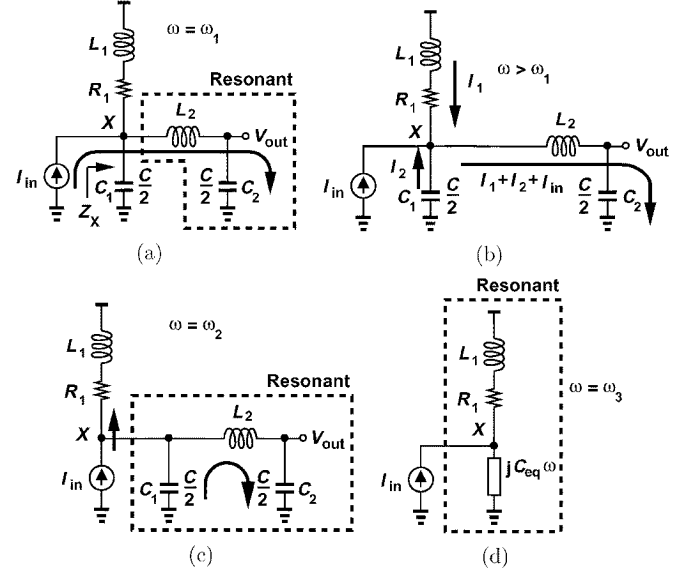


Fig. 5. Behavior of a triple-resonance circuit at different frequencies.

B. TRA Frequency Response

The series resonance of L_2 and C_2 depicted in Fig. 5(a) not only forces all of I_{in} to flow through C_2 , but reverses the sign of the impedance Z_X , thus making V_X negative for $\omega > \omega_1$. As illustrated in Fig. 5(b), I_1 and I_2 must therefore flow into node X and, together with I_{in} , pass through C_2 . The capacitive current I_2 multiplied by the impedance of C_2 creates a relatively constant output voltage as ω increases, while the inductive current I_2 introduces a roll-up in V_{out} .

Consequently, $|V_{out}/I_{in}|$ continues to rise until the π network consisting of C_1 , L_2 , and C_2 begins to resonate [Fig. 5(c)], presenting an infinite impedance at node X and allowing all of I_{in} to flow through R_1 and L_1 . This resonance frequency is given by

$$\omega_2 = \frac{1}{\sqrt{L_2 \frac{C_1 C_2}{C_1 + C_2}}} = \sqrt{2}\omega_1. \quad (5)$$

Since, at ω_2 , C_1 and C_2 carry equal and opposite currents

$$\begin{aligned} |V_{out}| &= |V_X| \\ &= |I_{in}| \sqrt{R_1^2 + L_1^2 \omega_2^2} \\ &= |I_{in}| \sqrt{\frac{3}{2}} R_1. \end{aligned} \quad (6)$$

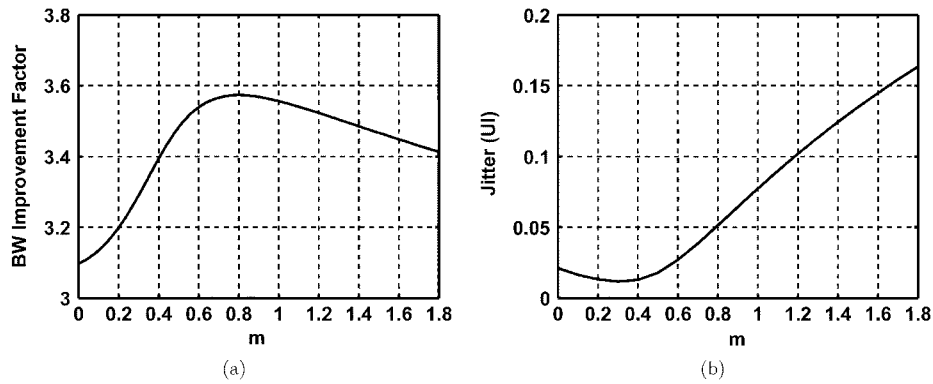


Fig. 6. (a) Bandwidth and (b) jitter of a triple-resonance stage plotted as a function of $m = L_2/L_1$.

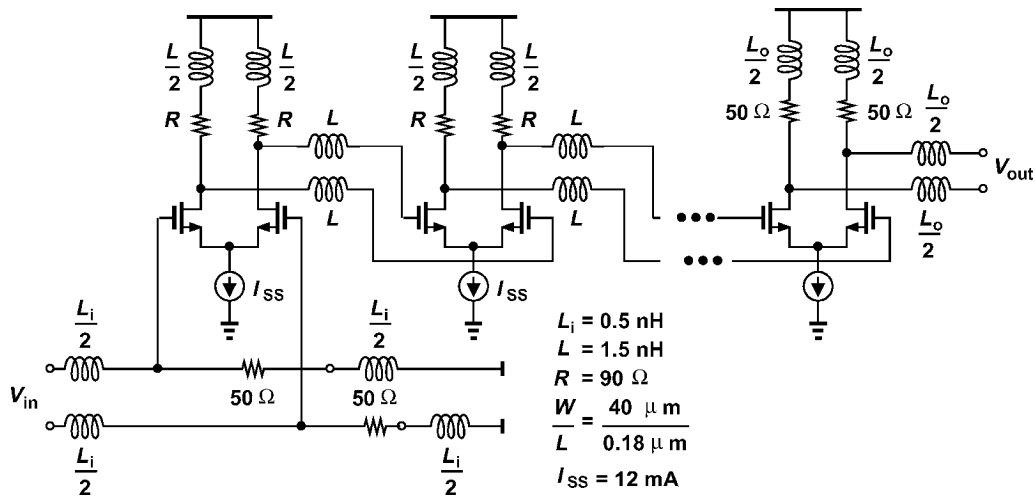


Fig. 7. Differential TRA.

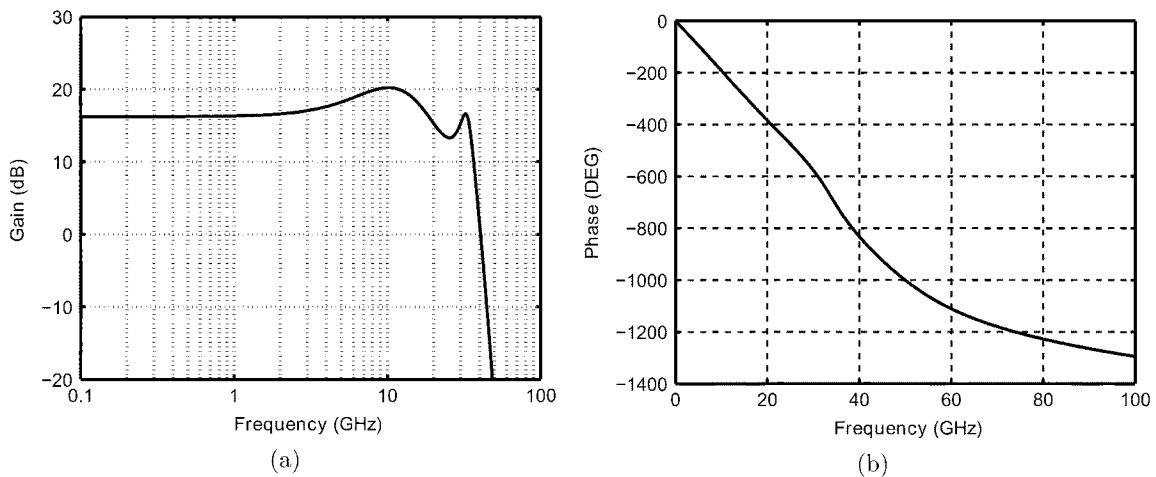


Fig. 8. TRA simulated (a) gain response and (b) phase response.

That is, the magnitude response of the amplifier exhibits a peaking of $\sqrt{3/2} \approx 1.8 \text{ dB}$.

For $\omega > \omega_2$, the π network becomes capacitive and $|V_{out}/I_{in}|$ begins to fall, returning to the midband value R_1 when the impedance of the π network resonates with L_1 [Fig. 5(d)]. This third resonance frequency is given by

$$\omega_3 = \sqrt[3]{6}\omega_1. \tag{7}$$

The -3-dB bandwidth exceeds this value and is approximately equal to

$$\begin{aligned} \omega_{-3\text{dB}} &\approx \sqrt{3}\omega_1 \\ &= \frac{2\sqrt{3}}{R_1 C}. \end{aligned} \tag{8}$$

In other words, the TRA improves the bandwidth of resistively loaded differential pairs by a factor of $2\sqrt{3} \approx 3.5$.

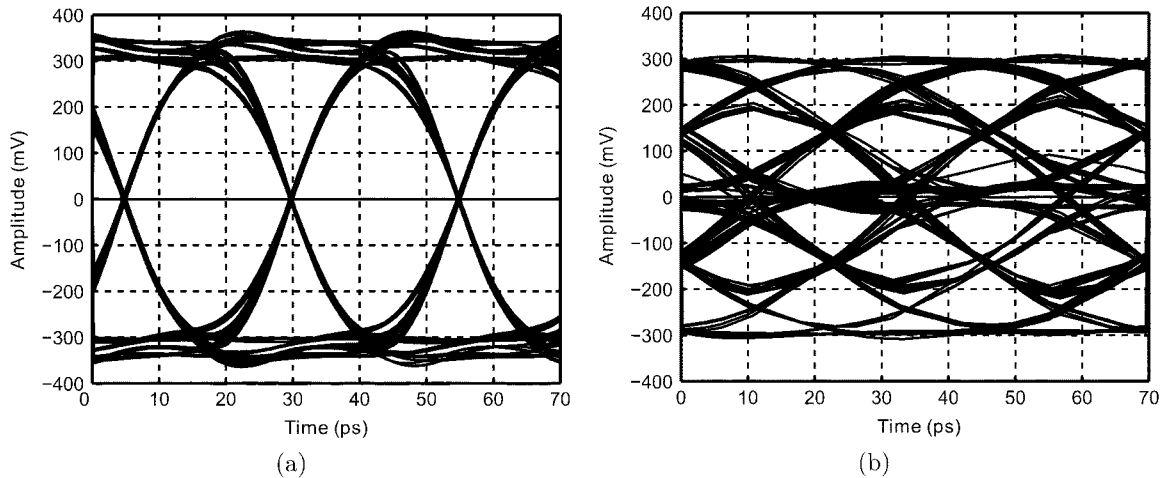


Fig. 9. Simulated 40-Gb/s output eye for (a) a TRA cascade and (b) a DA.

In our derivations, we have assumed $mL_1 = L_2$, where $m = 0.5$, but m can be increased to improve the bandwidth. The penalty is greater peaking and larger jitter. Fig. 6 plots the simulated bandwidth improvement factor and the amount of jitter of a triple-resonance stage as a function of m . As this ratio goes to 0.8, the bandwidth increases to some extent, but the jitter reaches 0.05 unit interval (UI). Thus, a ratio of 0.5 provides a reasonable compromise between bandwidth and jitter.

C. Amplifier Circuit

Fig. 7 depicts the overall 40-Gb/s amplifier. Five differential triple-resonance stages provide multiplicative gain, with each stage achieving a small-signal bandwidth of 32 GHz. With a loss of 5.3 dB in the last stage (due to a total load impedance of $25\ \Omega$), the overall gain reaches 15 dB. In the design process, we have observed that it is simpler to create a great internal gain and incur loss in the last stage than to make the last stage lossless.⁴

Fig. 8 plots the simulated frequency response of the TRA cascade. The amplifier has a midband gain of 15 dB and bandwidth of 26 GHz and exhibits a linear phase response up to 30 GHz. The total input-referred noise voltage is $0.4\ \text{mV}_{\text{rms}}$ from simulations.

The circuit of Fig. 7 exhibits several advantages over DAs. First, the load resistance of the internal stages need not be equal to $50\ \Omega$, allowing larger gain. Second, the series resistance of the inductors impacts the performance to a much lesser extent than in the transmission lines of DAs because it does not have a cumulative effect. Third, the voltage headroom constraints remain independent of the number of stages.

The 1.8-dB peaking illustrated in Fig. 4(c) is of concern in cascaded stages. However, the finite Q of the inductors lowers this effect considerably. As is evident from Fig. 8, the overall 40-Gb/s amplifier incurs a peaking of only 4 dB.

For the sake of comparison, a DA has been designed to have the same gain as the TRA amplifier. The DA design assumes a gain of 15 dB and hence a total g_m of 225 mS. With a total bias current of 40 mA, the undistributed transistor width

⁴For large-signal operation, the last stage must still provide reasonable swings. With a tail current of 30 mA, this stage can deliver a maximum swing of 1.5 V to $25\text{-}\Omega$ loads.

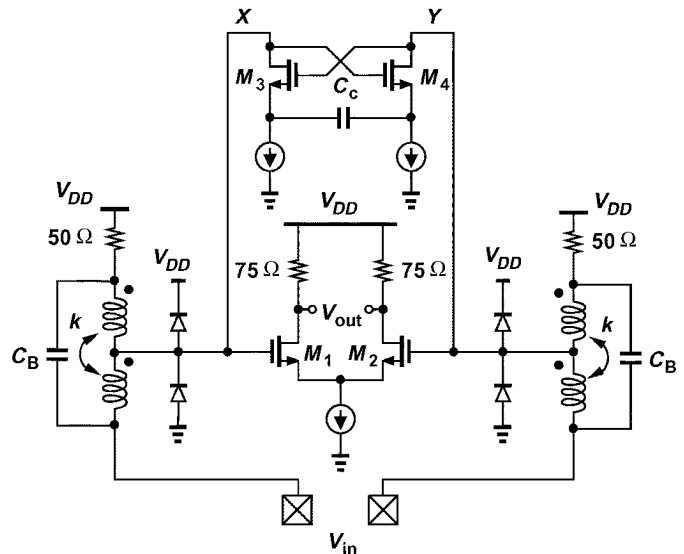


Fig. 10. Input ESD protection circuit.

necessary to achieve this value of g_m is equal to $288\ \mu\text{m}$, yielding $C_{\text{GS}} = 900\ \text{fF}$. Each transmission line is realized as a cascade of spiral inductors in metal 6, exhibiting (per nH of inductance) a series resistance of $10\ \Omega$ and a capacitance of $40\ \text{fF}$. Using simulations, it is determined that the transistor should be decomposed into five units and distributed over the lines with an inductance of $1.2\ \text{nH}$ per section. Fig. 9 plots the simulated output eye diagrams for both topologies, indicating that the proposed amplifier has a significantly wider bandwidth. The DA has a simulated small-signal bandwidth of 12 GHz.

IV. 40-Gb/s ESD CIRCUIT

As proposed in [7], T-coil networks can improve both the input matching and the bandwidth of ESD protection circuits. However, to approach 40 Gb/s, additional techniques are required. For a given ESD capacitance, losses in the T-coil still limit the bandwidth. This study describes two modifications of T-coil-based ESD circuits that extend the speed from 10

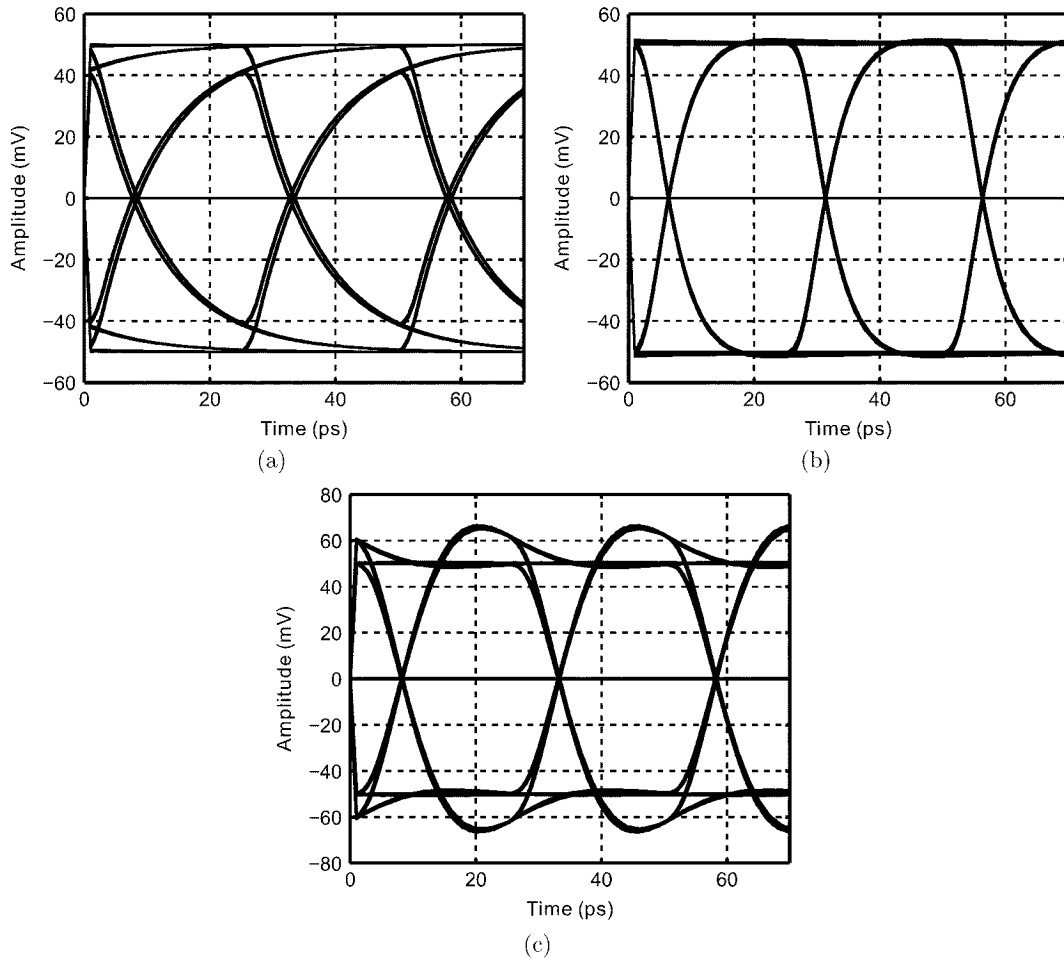


Fig. 11. ESD circuit output eye for (a) $C_c = C_B/10$, (b) $C_c = C_B/4$, and (c) $C_c = C_B$.

to 40 Gb/s with little compromise in voltage tolerance. Both concepts can be applied to output ESD protection circuits as well.

The first modification is to employ pn junctions rather than MOS-based topologies as ESD protection devices. A comparison of the results in [7] with those in this study suggests that pn junctions exhibit less capacitance for a given voltage tolerance.

The second modification is to lower the capacitance seen by the T-coil through the use of a negative impedance converter. As illustrated in Fig. 10, M_3 , M_4 , and C_c introduce a negative capacitance between nodes X and Y . As a result, the T-coils see much less load capacitance and themselves can be designed for a wider bandwidth.

The upper bound on the value of C_c is that which places the circuit at the edge of relaxation oscillation. This is quantified by scaling C_c with respect to the T-coil capacitance, C_B , if the transistor capacitances are negligible. Fig. 11 shows the simulated output eye of the ESD circuit with different values of C_c in response to an input PRBS pattern of $2^{23} - 1$. It is observed that $C_c = C_B/4$ leads to good performance. But, as C_c becomes comparable to the T-coil capacitance, the overshoot becomes significant. Therefore, the value of C_c can be one-fourth to one-half of the T-coil capacitance. For random data, C_c must remain within this bound to ensure minimal ringing and intersymbol interference (ISI).

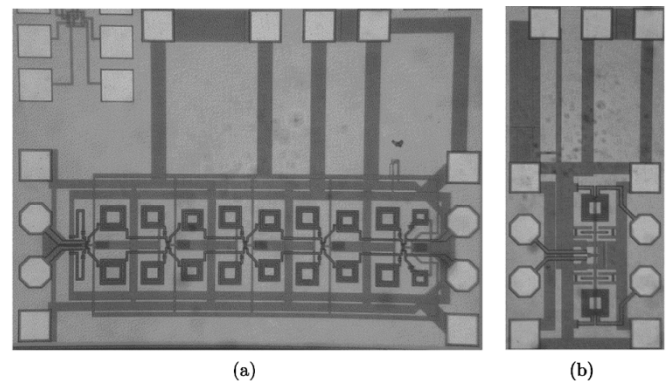
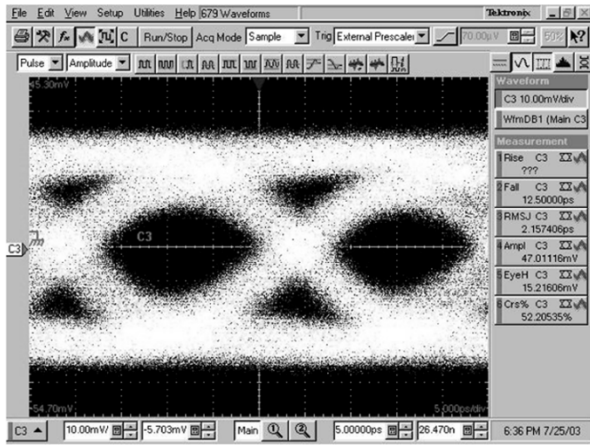


Fig. 12. Die photographs of (a) a TRA and (b) an ESD protection circuit.

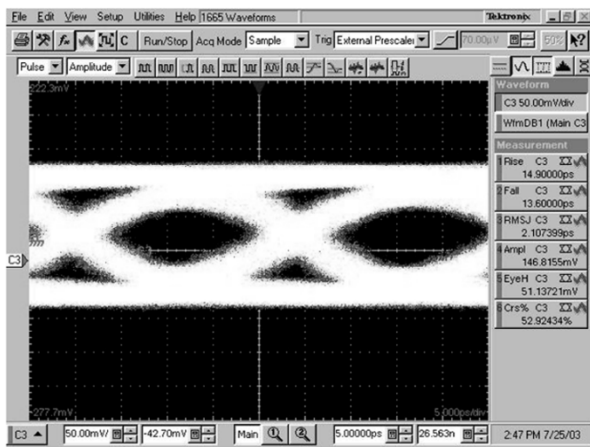
For large input swings, the cross-coupled pair becomes non-linear, providing less cancellation of capacitance. Nevertheless, simulations suggest that the output eye remains unchanged for a differential input swing as large as 500 mV.

V. EXPERIMENTAL RESULTS

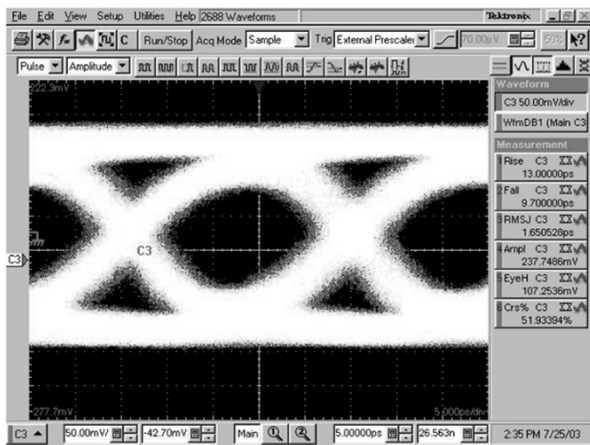
Both circuits have been fabricated in 0.18- μm CMOS technology and tested on a probe station using a 40-Gb/s bit stream generated by multiplexing four 10-Gb/s random data channels. The die photographs for both circuits are shown in Fig. 12.



(a)



(b)



(c)

Fig. 13. Measured amplifier single-ended output eye for an input signal level of (a) 20 mV_{pp}, (b) 50 mV_{pp}, and (c) 100 mV_{pp}. [Horizontal scale: 5 ps/div.; vertical scale: 10 mV/div. in (a), and 50 mV/div. in (b) and (c).]

Fig. 13 shows the single-ended output eyes of the amplifier for input levels of 20, 50, and 100 mV_{pp}, indicating a small-signal differential gain of 15 dB. Since these eyes are slightly less open than the simulation results, we decreased the bandwidth of each stage in simulations to obtain the same eye. This indicates that the small-signal bandwidth is about 22 GHz.⁵ The

⁵The reduced bandwidth is attributed to two sources: higher load resistance values in each stage and the loss of connectors, cables, and probes in the setup.

TABLE I
PERFORMANCE COMPARISON OF A TRA WITH PRIOR ART

| Design | Technology | Gain (dB) | BW (GHz) | A _v × BW (GHz) | V _{DD} (V) | Power (mW) |
|------------------|---|-----------|-----------|---------------------------|---------------------|------------|
| [8] | 0.5 μm SOS MOS | 5 | 10 | 17.8 | - | - |
| [9] | 0.18 μm CMOS | 50 | 9.4 | 2970 | 1.8 | 150 |
| [10] | 0.18 μm CMOS | 10 | 10 | 31.6 | - | - |
| [11] | 0.18 μm CMOS | 10.6 | 14 | 47.4 | 1.3 | 52 |
| This Work | 0.18 μm CMOS | 15 | 22 | 124 | 2.2 | 190 |

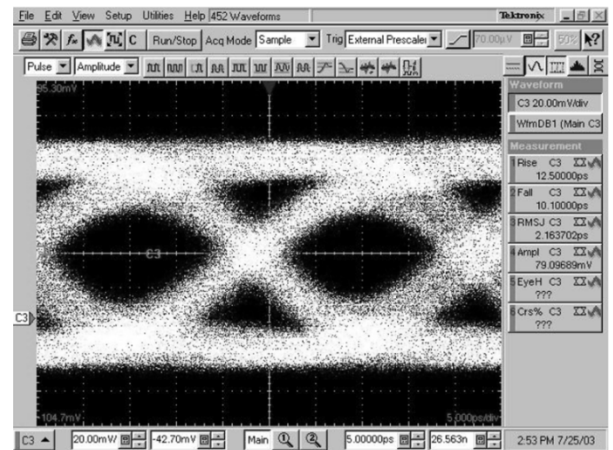


Fig. 14. Measured output eye of the ESD circuit. [Horizontal scale: 5 ps/div.; vertical scale: 20 mV/div.].

direct measurement of the bandwidth is not possible because network analyzers handle only single-ended signals. The circuit consumes 190 mW from a 2.2-V supply.

Table I compares the performance of the amplifier with that of recent work in CMOS technology.⁶ (Even with a 2.2-V supply, the gate–drain voltage of the transistors does not exceed 1.2 V.) The proposed circuit achieves substantially larger bandwidth and gain–bandwidth product than DAs. Also, a comparison of [9] (cascaded stages) and [10] reveals that practical DAs achieve a much lower gain–bandwidth product than cascaded stages. None of the circuits in [8], [10], and [11] have been tested with random data to reveal effects such as ringing or ISI.

Fig. 14(a) shows the measured single-ended 40-Gb/s output eye of the ESD circuit.⁷ For four samples, the human-body model tolerance is 700–800 V while the machine-model tolerance is 100 V.

⁶The results reported in [11] are based on power gain whereas this table calculates the GBW in terms of voltage gain. The comparison remains consistent and fair in either case.

⁷In practice, a larger bandwidth is necessary to accommodate the bandwidth limitations of the subsequent stages.

VI. CONCLUSION

A TRA that allows CMOS technology to operate at 40 Gb/s is presented. The proposed architecture outperforms DAs by at least a factor of two in terms of gain–bandwidth product. In addition, a modified T-coil-based ESD circuit that uses pn junctions and negative impedance converters lowers the parasitic ESD capacitance and extends the speed from 10 to 40 Gb/s.

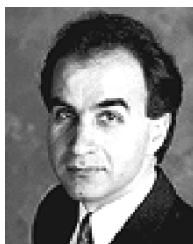
REFERENCES

- [1] H. Tao, D. K. Shaeffer, X. Min, S. Benyamin, V. Conditto, S. Kudszus, L. Qinghung, A. Ong, A. Shahani, S. Xiaomin, W. Wong, and M. Tarsia, "40–43-Gb/s OC-768 16:1 MUX/CMU chipset with SFI-5 compliance," *IEEE J. Solid-State Circuits*, vol. 38, pp. 2169–2180, Dec. 2003.
- [2] A. Ong, S. Benyamin, J. Cancio, V. Conditto, T. Labrie, L. Qinghung, J. P. Mattia, D. K. Shaeffer, A. Shahani, S. Xiaomin, H. Tao, M. Tarsia, W. Wong, and X. Min, "A 40–43-Gb/s clock and data recovery IC with integrated SFI-5 1:16 demultiplexer in SiGe technology," *IEEE J. Solid-State Circuits*, vol. 38, pp. 2155–2168, Dec. 2003.
- [3] G. Freeman, M. Meghelli, Y. Kwark, S. Zier, A. Rylyakov, M. A. Sorna, T. Tanji, O. M. Schreiber, K. Walter, R. Jae-Sung, B. Jagannathan, A. Joseph, and S. Subbanna, "40-Gb/s circuits built from a 120-GHz f_T SiGe technology," *IEEE J. Solid-State Circuits*, vol. 37, pp. 1106–1114, Sept. 2002.
- [4] B. Razavi, *Design of Integrated Circuits for Optical Communications*. New York: McGraw-Hill, 2003.
- [5] T. T. Y. Wong, *Fundamentals of Distributed Amplification*. Norwood, MA: Artech House, Oct. 1993.
- [6] E. A. Henry, "Practical design of video amplifiers," *QST*, pp. 32–38, May 1945.
- [7] S. Galal and B. Razavi, "Broadband ESD protection circuits in CMOS technology," in *ISSCC Dig. Tech. Papers*, Feb. 2003, pp. 182–183.
- [8] P. F. Chen, R. A. Johnson, M. Wetzel, P. R. de la Houssaye, G. A. Garcia, P. M. Asbeck, and I. Lagnado, "Silicon-on-sapphire MOSFET distributed amplifier with coplanar waveguide matching," in *IEEE RFIC Symp. Dig. Tech. Papers*, 1998, pp. 161–164.
- [9] S. Galal and B. Razavi, "10-Gb/s limiting amplifier and laser/modulator driver in 0.18- μ m CMOS technology," in *ISSCC Dig. Tech. Papers*, Feb. 2003, pp. 188–189.
- [10] B. M. Frank, A. P. Freundorfer, and Y. M. Antar, "Performance of 1–10 GHz traveling wave amplifiers in 0.18- μ m CMOS," in *IEEE MWCL Dig. Tech. Papers*, vol. 12, Sept. 2002, pp. 327–329.
- [11] R.-C. Liu, C.-S. Lin, K.-L. Deng, and H. Wang, "A 0.5–14-GHz 10.6-dB CMOS cascode distributed amplifier," in *IEEE Symp. VLSI Circuits Dig. Tech. Papers*, 2003, pp. 139–140.



Sherif Galal received the B.S. degree in electrical engineering and the M.S. degree from Ain Shams University, Cairo, Egypt, in 1994 and 1999, respectively. He is currently working toward the Ph.D. degree at the University of California, Los Angeles (UCLA).

Since September 1999, he has been with UCLA, where his research focuses on high-speed circuits for broadband and RF applications.



Behzad Razavi (S'87–M'90–SM'00–F'03) received the B.Sc. degree in electrical engineering from Sharif University of Technology, Tehran, Iran, in 1985 and the M.Sc. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1988 and 1992, respectively.

He was an Adjunct Professor at Princeton University, Princeton, NJ, from 1992 to 1994, and at Stanford University, Stanford, CA, in 1995. He was with AT&T Bell Laboratories and Hewlett-Packard Laboratories until 1996. Since 1996, he has been Associate Professor and subsequently Professor of electrical engineering at the University of California, Los Angeles. He is the author of *Principles of Data Conversion System Design* (IEEE Press, 1995), *RF Microelectronics* (Prentice Hall, 1998) (also translated into Japanese), *Design of Analog CMOS Integrated Circuits* (McGraw-Hill, 2001) (also translated into Chinese and Japanese), and *Design of Integrated Circuits for Optical Communications* (McGraw-Hill, 2003), and the editor of *Monolithic Phase-Locked Loops and Clock Recovery Circuits* (IEEE Press, 1996), and *Phase-Locking in High-Performance Systems* (IEEE Press, 2003). His current research includes wireless transceivers, frequency synthesizers, phase locking and clock recovery for high-speed data communications, and data converters.

Dr. Razavi received the Beatrice Winner Award for Editorial Excellence at the 1994 IEEE International Solid-State Circuits Conference (ISSCC), the best paper award at the 1994 European Solid-State Circuits Conference, the Best Panel Award at the 1995 and 1997 ISSCC, the TRW Innovative Teaching Award in 1997, and the Best Paper Award at the IEEE Custom Integrated Circuits Conference in 1998. He was the co-recipient of both the Jack Kilby Outstanding Student Paper Award and the Beatrice Winner Award for Editorial Excellence at the 2001 ISSCC. He has been recognized as one of the top ten authors in the 50-year history of ISSCC. He served on the Technical Program Committees of the ISSCC from 1993 to 2002 and the VLSI Circuits Symposium from 1998 to 2002. He has also served as Guest Editor and Associate Editor of the IEEE JOURNAL OF SOLID-STATE CIRCUITS, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS, and the *International Journal of High Speed Electronics*. He is an IEEE Distinguished Lecturer.