

# Prospects of CMOS Technology for High-Speed Optical Communication Circuits

Behzad Razavi

Electrical Engineering Department, University of California, Los Angeles, CA 90095  
and Transpectrum Technologies, Los Angeles, CA 90024

## Abstract

**This paper explores the potential of CMOS technology for circuits operating at tens of gigahertz in an optical communications environment. An overview of modern CMOS processes is given and a generic optical system illustrating integration challenges is studied. The design of high-speed building blocks such as amplifiers, oscillators, and phase detectors is also described.**

## I. INTRODUCTION

The explosive demand for high data rates has revitalized optical communications, motivating intensive research on high-speed devices, circuits, and systems. The new optical revolution is reminiscent of the monumental change that the RF design paradigm began to experience in the early 1990s: modular, general-purpose building blocks are gradually replaced by end-to-end solutions that benefit from device/circuit/architecture codesign, and mainstream VLSI technologies such as Bi-CMOS and CMOS continue to take over the territories thus far claimed by GaAs and InP devices.

This paper examines the potential of CMOS technology for high-speed optical communications. We begin with an overview of modern CMOS processes, justifying their use for high-speed design. Next, we study a generic optical system and describe the speed and noise issues in the integration of transceivers. We then present the design of important building blocks such as amplifiers, oscillators, and phase detectors in CMOS technology. As a framework, we use the OC-768 standard for the design targets in the paper.

## II. WHY CMOS?

### A. General Attributes

Aggressive scaling and the competition to follow Moore's Law have improved the intrinsic speed of MOSFETs by more than three orders of magnitude in the past 30 years. The  $f_T$  of NMOS transistors in the 0.15- $\mu\text{m}$  and 0.13- $\mu\text{m}$  generations approaches 80 GHz and is likely to exceed 120 GHz for 0.1- $\mu\text{m}$  devices. Also, more relevant benchmarks such as differential ring oscillators and frequency dividers exhibit maximum operating speeds of several tens of gigahertz in the 0.13- $\mu\text{m}$  generation.

Despite rapid scaling, the present MOS technologies may appear inadequate for systems operating in the range of 40 Gb/s.

However, another important property of CMOS processes, namely, the availability of multitude of metal layers, can substantially boost the performance by providing high-quality passive devices. As noted throughout this paper, monolithic components such as inductors, transmission lines, MOS varactors, and linear capacitors prove essential to extending the capabilities of CMOS technology to 40 Gb/s. Fortunately, the extensive research carried out on passive devices in both RF IC design and millimeter-wave circuit design can be exploited in optical systems as well.

In addition to scaling the dimensions and providing many metal layers, CMOS technology exhibits two other attributes germane to circuit design for optical communications. First, the inevitable scaling of the supply voltage does reduce the overall power dissipation of the system even though it creates many difficulties in the design of the building blocks. For example, a 1-to-16 demultiplexer (DMUX) with low-voltage differential signaling (LVDS) outputs across 100- $\Omega$  differential loads typically draws a supply current of  $16 \times 5 \text{ mA} = 80 \text{ mA}$ , a significant fraction of the overall transceiver's current. Thus, if the supply voltage is decreased from 3 V to 1.5 V, the DMUX power dissipation drops considerably.

The second attribute relates to the cost. Owing to lower fabrication cost, higher yield, and greater density of MOS devices, CMOS implementations prove more economical than their BiCMOS or III-V counterparts. While the cost advantage may not be apparent for low-complexity circuits such as transimpedance amplifiers (TIAs) and limiters, it does rise as a distinguishing factor when a full transceiver must be integrated on a single chip. In systems where many channels are carried on different wavelengths or on a bundle of fibers, multiple transceivers must be realized monolithically, further underlining the potential of CMOS technology. Moreover, the shift of paradigm towards integrating transceivers and framers on the same chip may select CMOS technology as the only viable solution. This trend is similar to the increasing sophistication that has appeared in RF CMOS transceivers.

### B. Passive Devices

The extensive work on inductors, varactors, and oscillators inherited from RF CMOS design proves invaluable in high-speed applications as well. Since spiral inductors occupy a large area, stacked structures connected in series can be used [Fig. 1(a)]. Also, the bottom spiral can be moved away from the top one to reduce the parasitic capacitance significantly

[Fig. 1(b)] [1]. Addition of a grounded  $n$ -well shield under the inductors suppresses cross-talk, an important feature in integrating various broadband circuits on the same substrate.

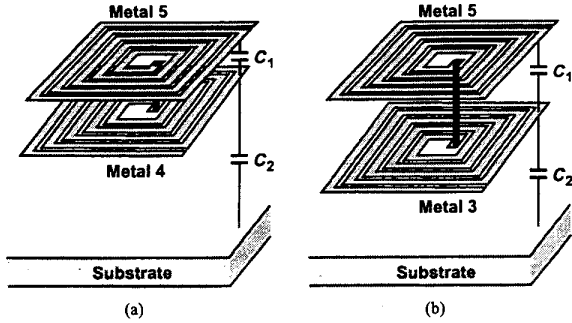


Fig. 1. (a) Stacked inductors, (b) modification to lower the total capacitance.

At speeds of tens of gigahertz, the use of on-chip transmission lines becomes an attractive technique. Shown in Fig. 2 are two topologies suited to integration in CMOS processes. The coplanar structure in Fig. 2(a) provides a relatively small capacitance per unit length but it suffers from resistive loss in the substrate. The microstrip line depicted in Fig. 2(b) exhibits a slightly greater capacitance but terminates the electric field lines on metal 1, thereby lowering the loss. These structures have been successfully used in CMOS circuits operating at tens of gigahertz.

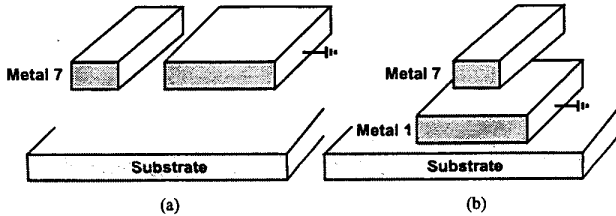


Fig. 2. (a) Coplanar transmission line, (b) microstrip line.

MOS varactors (Fig. 3) directly benefit from the scaling of the channel length as their quality factor is determined by the  $n$ -well resistance between the source and drain terminals. Offering a wider range and accommodating both positive and negative voltages, MOS varactors provide much more flexibility in design than pn junctions do.

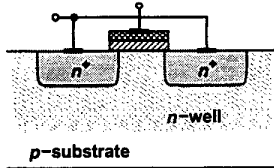


Fig. 3. MOS varactor.

At low supply voltages, capacitive coupling between cascaded stages may relax the voltage headroom constraints. However, both the bottom-plate parasitic capacitance and the low density of typical "native" capacitor structures make their use difficult. A practical solution is the "fringe" capacitor shown in Fig. 4, whereby the large fringe capacitance between

adjacent metal lines is heavily exploited. With six or seven metal layers, a bottom-plate parasitic of only a few percent and a density of about  $0.5 \text{ fF}/\mu\text{m}$  can be achieved.

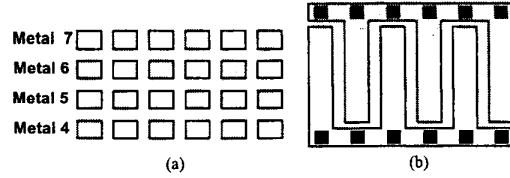


Fig. 4. Fringe capacitor, (a) cross section, (b) top view.

### III. SYSTEM OVERVIEW

Figure 5 shows a typical optical system. In the transmitter (TX), a number of channels are multiplexed into a high-speed

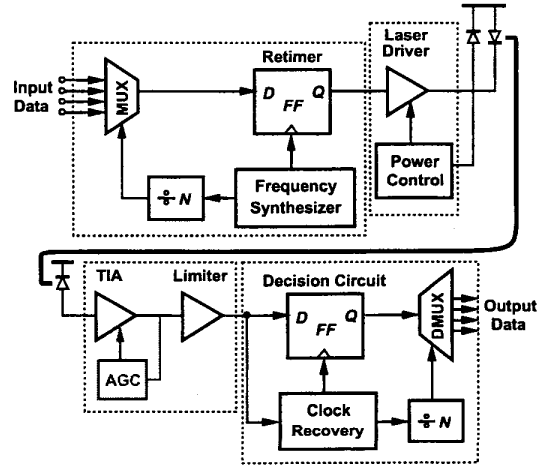


Fig. 5. Optical transceiver system.

data stream, the result is retimed and applied to a laser driver, and the optical output thus produced is delivered to the fiber. A frequency synthesizer generates clocks for both the multiplexer (MUX) and the retiming flipflop (FF). Also, since the laser output power varies with temperature and aging, a monitor photodiode (PD) and a power control circuit continuously adjust the output level of the driver.

In the receiver (RX), a photodiode converts the optical signal to a current and a transimpedance amplifier and a limiter raise the signal swing to logical levels. [The TIA may incorporate automatic gain control (AGC) to accommodate a wide range of input currents.] Subsequently, a clock recovery circuit extracts the clock from the data with proper edge alignment and retimes the data by a "decision circuit." The result is then demultiplexed, thereby producing the original channels.

The transmitter of Fig. 5 entails several issues that manifest themselves at high speeds and/or in scaled IC technologies. Since the jitter of the transmitted data is determined by primarily that of the synthesizer, a robust, low-noise phase-locked loop (PLL) with high supply and substrate rejection becomes

essential. Furthermore, the design of skew-free, synchronous multiplexers proves difficult at high data rates.

Another critical challenge arises from the laser driver, a circuit that must deliver tens of milliamperes of current with very short rise and fall times. Since laser diodes may experience large voltage swings between on and off states, the driver design becomes more difficult as scaled technologies impose lower supply voltages. The package parasitics also severely limit the speed with which such high currents can be switched to the laser.

The receiver of Fig. 5 also presents many problems. The noise, gain, and bandwidth of the TIA and the limiter directly impact the sensitivity and speed of the overall system, raising additional issues as the supply voltage scales down. Moreover, the clock and data recovery functions must provide a high speed, tolerate long runs (sequences of identical bits), and satisfy stringent jitter and bandwidth requirements.

Full integration of the transceiver shown in Fig. 5 on a single chip raises a number of concerns. The high-speed digital signals in the MUX and DMUX may corrupt the receiver input or the oscillators used in the synthesizer and the CDR circuit. The high slew rates produced by the laser driver may lead to similar corruptions and also desensitize the TIA. Finally, since the VCOs in the transmit synthesizer and the receive CDR circuit operate at slightly different frequencies (with the difference given by the mismatch between the crystal frequencies in two communicating transceivers), they may pull each other, generating substantial jitter.

The above issues have resulted in multichip solutions that integrate the noisy and sensitive functions on different substrates. The dashed boxes in Fig. 5 indicate a typical partitioning, suggesting the following single-chip blocks: the synthesizer/MUX circuit (also called the “serializer”), the laser driver along with its power control circuitry, the TIA/limiter combination, and the CDR/MUX circuit (also called the “deserializer”). Recent work has integrated the serializer and deserializer (producing a “serdes”) but the TX and RX amplifiers remain in isolation.

## IV. BUILDING BLOCKS

### A. Broadband Amplification

An attractive solution for low-voltage broadband amplifiers is inductive peaking. Owing to the extensive work on monolithic inductors in RF design, this method can now be realized with accurate modeling and prediction of the performance in optical communication circuits as well. Interestingly, inductor quality factors ( $Q$ 's) as low as 3 to 4 prove adequate for increasing the bandwidth, allowing the use of simple, compact spiral structures.

Figure 6(a) shows a differential stage incorporating inductive peaking. It can be shown that an ideal inductor increases the bandwidth by approximately 82% if a 7.5% overshoot in the step response is acceptable. With the finite  $Q$  and parasitic capacitance of the inductors included, the enhancement is around 50%, still quite a significant factor.

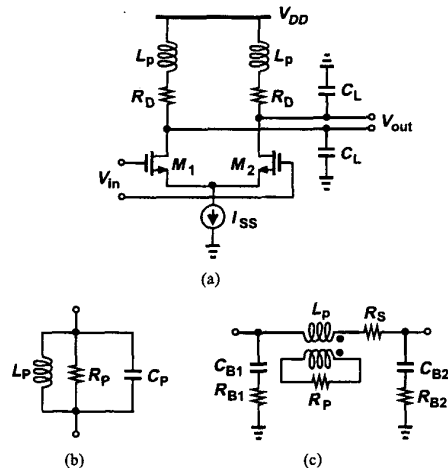


Fig. 6. (a) Inductive peaking, (b) simple inductor model, (c) more complete inductor model.

An interesting difficulty in modeling the inductors in the above circuit arises from the narrowband nature of the definition of the  $Q$ , an issue rarely encountered in RF design. Figure 6(b) depicts a rough model where  $R_P/(L_P\omega)$  yields the correct  $Q$  at about 3/4 of the  $-3$ -dB bandwidth. The approximation is reasonable because the inductor manifests itself only near the high end of the band. Alternatively, a more complete model such as that in Fig. 6(c) can be used. Here,  $R_S$  denotes the effective series resistance,  $R_{B1}$  and  $R_{B2}$  represent the resistance seen by the electric coupling to the substrate,  $R_P$  models the resistance seen by the magnetic coupling to the substrate, and the capacitors approximate the parasitic capacitances. While the values of some of the components in this model do vary with frequency, the overall model can be fitted to measured data over a broader range than the parallel tank of Fig. 6(b) can.

Another technique suited to broadband signals is distributed amplification. Illustrated in Fig. 7 in differential form, such an amplifier distributes the transistor capacitances along a trans-

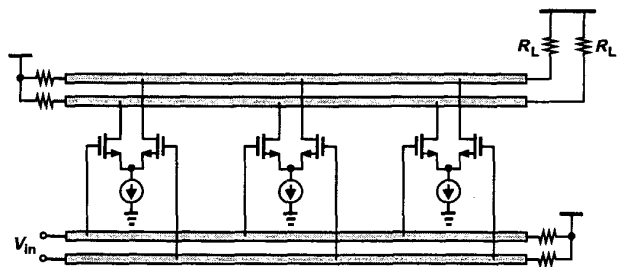


Fig. 7. Differential distributed amplifier.

mission line, thereby allowing a greater gain as more sections are added and hence relaxing the trade-off between gain and bandwidth. Since transmission lines in CMOS technology exhibit a moderate loss, a relatively high gain can be achieved. In fact, with a large number of sections drawing currents from the termination resistors, the overall gain may be limited by

the voltage headroom rather than the line loss. Examples of CMOS distributed amplifiers are reported in [2, 3].

### B. Oscillators and Frequency Dividers

The transmitter of Fig. 5 requires that the last retiming flipflop be driven by a full-rate clock, e.g., 40 GHz for OC-768. Note that if the flipflop is omitted or the 40-GHz clock is derived by doubling the frequency of a 20-GHz signal, then mismatches yield considerable jitter in the transmitted data. For such speeds, an LC oscillator may be used (Fig. 8). Monolithic

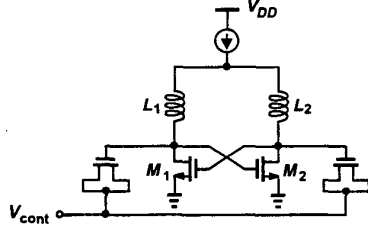


Fig. 8. LC VCO using MOS varactors.

inductors and varactors provide sufficiently high  $Q$ 's at 40 GHz to afford low-noise oscillation. Alternatively, placing the distributed amplifier of Fig. 7 in a negative-feedback loop leads to an oscillator [2].

The frequency dividers employed in the transmit PLL must operate at high speeds while driving the large capacitance of the chain of multiplexers. For example, the divider following the VCO must typically drive nine differential pairs while sensing a 40-GHz clock. A candidate for division at high speeds is an injection-locked oscillator running at half the input frequency (Fig. 9). Addition of tuning can widen the frequency range of

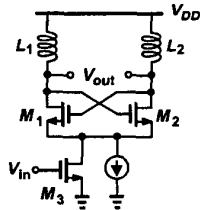


Fig. 9. Injection-locked frequency divider.

injection, allowing for process and temperature variations.

### C. Half-Rate Phase Detectors

If the data rate is higher than the maximum tolerable speed of phase detectors and VCOs, a half-rate CDR architecture can be used. The idea is to run the VCO at a frequency equal to half of the data rate, thereby relaxing the design of the circuits in the signal path. Half-rate architectures usually demultiplex the data as well.

The principal challenge in half-rate CDR circuits relates to the design of phase and frequency detectors that operate properly while sensing full-rate data and a half-rate clock. Figure 10 depicts an example of a linear half-rate PD [4]. The circuit employs four D latches and two XOR gates. Since latches  $L_1$  and  $L_2$  sample  $D_{in}$  on rising and falling edges

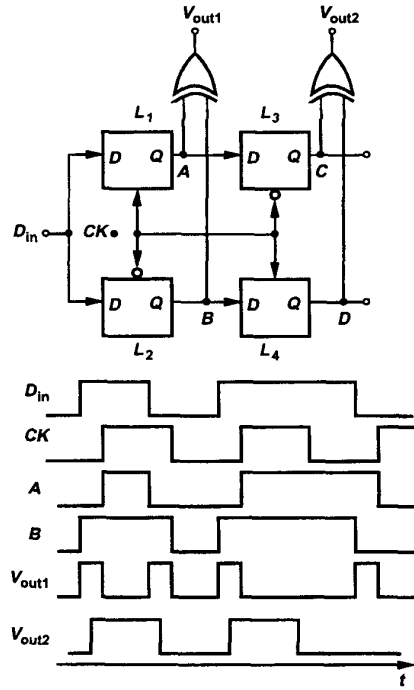


Fig. 10. Half-rate linear PD.

of  $CK$ ,  $A \oplus B$  produces a pulse each time a data transition occurs between a rising edge and a falling edge of the half-rate clock. The waveforms at  $C$  and  $D$  are identical except for a phase difference equal to half of the clock period. Thus,  $C \oplus D$  produces a constant-width pulse on every data transition, serving as a reference. Note that the waveforms at  $C$  and  $D$  are the retimed, demultiplexed versions of the input stream. Thus, no explicit retiming of data is required.

Other examples of half-rate phase and frequency detectors are described in [5, 6].

### REFERENCES

- [1] A. Zolfaghari, A. Y. Chan, and B. Razavi, "Stacked Inductors and Transformers in CMOS Technology," *IEEE Journal of Solid-State Circuits*, vol. 36, pp. 620-628, April 2001.
- [2] B. Kleveland et al, "Monolithic CMOS Distributed Amplifier and Oscillator," *ISSCC Dig. of Tech. Papers*, pp. 70-71, Feb. 1999.
- [3] B. Ballweber, R. Gupta, D. Allstot, "Fully-Integrated CMOS RF Amplifiers," *ISSCC Dig. of Tech. Papers*, pp. 72-73, Feb. 1999.
- [4] J. Savoj and B. Razavi, "A 10-Gb/s CMOS Clock and Data Recovery Circuit with a Half-Rate Linear Phase Detector," *IEEE Journal of Solid-State Circuits*, vol. 36, pp. 761-768, May 2001.
- [5] M. Wurzer et al, "A 40-Gb/s Integrated Clock and Data Recovery Circuit in a 50-GHz  $f_T$  Silicon Bipolar Technology," *IEEE Journal of Solid-State Circuits*, vol. 34, pp. 1320-1324, Sept. 1999.
- [6] J. Savoj and B. Razavi, "A 10-Gb/s CMOS Clock and Data Recovery Circuit with Frequency Detection," *ISSCC Dig. of Tech. Papers*, pp. 78-79, Feb. 2001.