Prospects of CMOS Technology for High-Speed Optical Communication Circuits

Behzad Razavi, *Senior Member, IEEE*

Invited Paper

***Abstract*—This paper describes the capabilities of deep-submicron CMOS technologies for the realization of highly integrated optical communication transceivers in the range of tens of gigabits per second. Following an overview of a CMOS process, the design of traditional and modern transceivers is presented and speed and integration issues are discussed. Next, the problem of equalization is addressed. Finally, the design of critical building blocks such as broadband amplifiers and high-speed oscillators is described and a method of estimating the jitter is introduced.**

***Index Terms*—Broad-band circuits, clock recovery, CMOS transceivers, inductive peaking, jitter, optical communication, oscillators.**

I. Introduction

THE rapidly growing volume of data in telecommunication networks has rekindled interest in high-speed optical and electronic devices and systems. This new wave entails three important trends similar to those which the radio-frequency (RF) design paradigm began to experience in the early 1990s.

1) Modular, general-purpose building blocks are gradually replaced by end-to-end solutions that benefit from device/circuit/architecture codesign.
2) Greater levels of integration on a single chip provide higher performance and lower cost.
3) Mainstream VLSI technologies such as CMOS and BiCMOS continue to take over the territories thus far claimed by GaAs and InP devices.

This paper describes the prospects of deep-submicron CMOS technologies for optical systems and circuits operating at high speeds with high functional complexity. As a framework, the paper aims to quantify the capabilities and limitations of CMOS processes for 40-Gb/s systems while extending well-known high-speed bipolar techniques [1] and RF design concepts to CMOS realizations.

Section II provides a brief overview of a typical CMOS process and some of its benchmarks that are relevant to optical communication (OC) circuits. Section III describes a traditional optical transceiver (TRX) system, identifying speed, noise, and integration issues. Section IV examines a modern optical transceiver, presenting the anticipated complexity and, hence, the need for CMOS technology. Section V deals with the design of building blocks and Section VI summarizes the advantages of CMOS technology for optical systems.

Manuscript received January 1, 2002; revised April 22, 2002.

The author is with the Department of Electrical Engineering, University of California, Los Angeles, CA 90095 USA (e-mail: razavi@ee.ucla.edu).

Publisher Item Identifier 10.1109/JSSC.2002.801195.

II. Attributes of CMOS Technology

Aggressive scaling resulting from the competition to follow Moore's Law has improved the intrinsic speed of MOSFETs by more than three orders of magnitude in the past 30 years. Fig. 1 illustrates a circuit designer's view of a modern CMOS process, e.g., a 0.13-μm generation. In addition to nMOS and pMOS transistors, the technology provides:

1) a deep n-well, which can be utilized to reduce substrate noise coupling;
2) a MOS varactor, which can serve in voltage-controlled oscillators (VCOs);
3) eight layers of metal, M1–M8, which can form many useful structures such as inductors, capacitors, and transmission lines.

A. Active Devices

In order to quantify the "raw" capabilities of the technology, we study a number of relevant benchmarks. Fig. 2 plots the simulated f_T of nMOS and pMOS transistors as a function of the gate–source overdrive voltage, $V_{\rm GS} - V_{\rm TH}$, for various process and temperature conditions. The f_T of nMOS devices falls to about 62 GHz at the slow high-temperature corner, suggesting difficulties in operating at 40 GHz.

A more realistic benchmark for OC circuits is the maximum speed of differential ring oscillators. Simulations using 0.13-μm devices indicate that three-stage differential rings with resistive loads oscillate at about 18 GHz. Since in an OC transceiver, such an oscillator drives a large number of latches, its output buffer must employ high currents and wide transistors, further lowering the speed.

Another important benchmark for OC circuits is the performance of divide-by-two circuits. Fig. 3 plots the maximum required (differential) clock swing for a static current-steering flip-flop divider as a function of the clock frequency (obtained by simulations). In an OC transceiver, such a divider must drive at least another divider and a 2-to-1 multiplexer (MUX), exhibiting a lower speed.

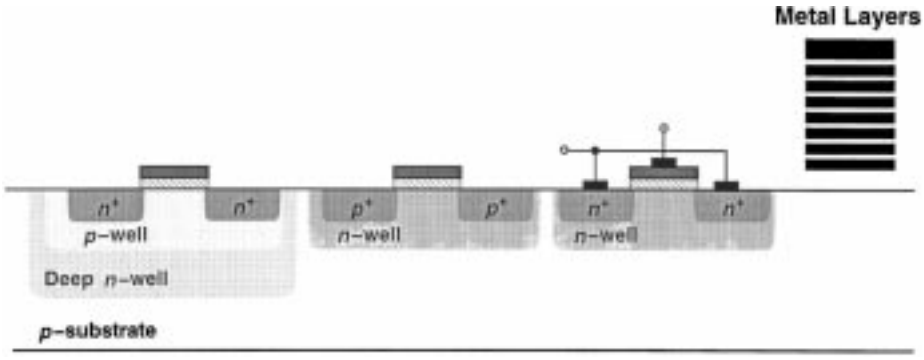

Fig. 1. Typical CMOS process.

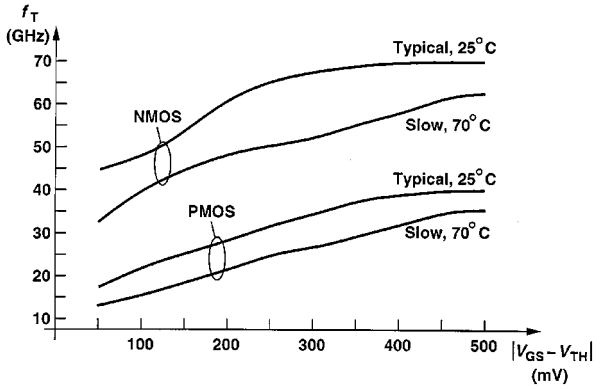

Fig. 2. Transit frequency of 0.13-μm nMOS and pMOS devices.

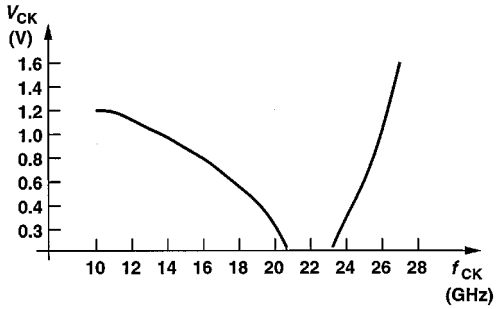

Fig. 3. Divide-by-two circuit sensitivity.

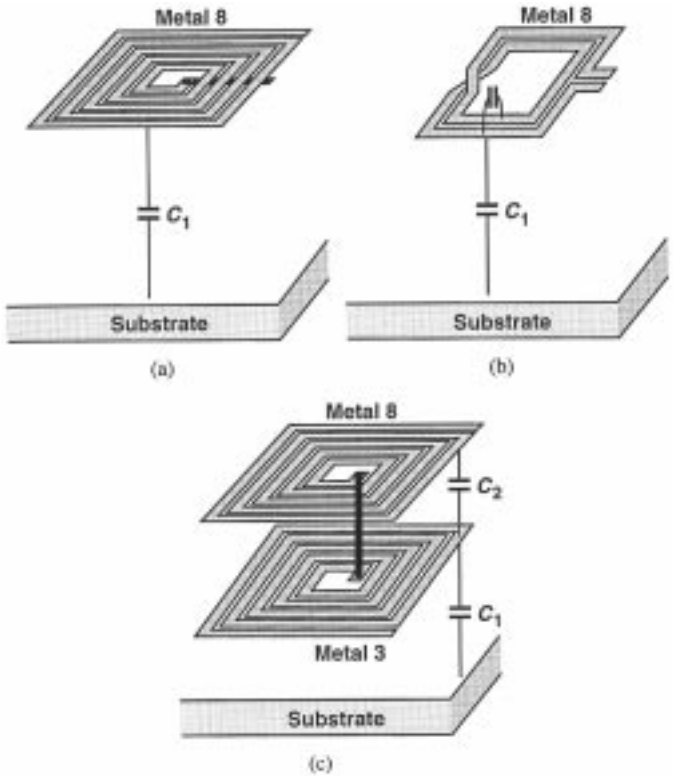

Fig. 4. Spiral inductor structures.

B. Passive Devices

The above speed limitations of CMOS technology can be overcome through the addition of passive components to the circuit designer's device library. Examples include spiral inductors, transmission lines (T lines), and MOS varactors. The process back-end illustrated in Fig. 1 reveals the availability of eight metal layers in recent generations of CMOS technology, providing fertile grounds for new device structures.

Inductors: Extensive studies of monolithic inductors in the context of RF design have created structures with acceptable quality factors (Qs), high self-resonance frequencies, and moderate dimensions. We are fortunate to have inherited the vast body of knowledge on inductors. For optical communication circuits, three types of inductor structures prove useful (Fig. 4). Measurements indicate a Q of 5 to 6 for the single-ended spiral of Fig. 4(a) in the frequency range of 30–40 GHz and a Q of 10–11 for the differential structure of Fig. 4(b) in the range of 15–30 GHz.[1] Since the interwinding capacitance sustains a much greater voltage in the differential inductor than in the single-ended spiral, the latter topology achieves a higher self-resonance frequency $f_{\rm SR}$.

In multiphase oscillators or inductively peaked broadband amplifiers, a multitude of inductors are necessary, mandating a small area for each. A candidate for this purpose is the stacked structure shown in Fig. 4(c), where several spirals are placed in series. The strong mutual coupling between every two spirals yields a total inductance of roughly n^2L_u, where n denotes the number of layers and L_u the inductance of one spiral. With eight metal layers available, various topologies can be envisaged that achieve a high $f_{\rm SR}$, high inductance, and/or high Q. Fig. 5(a) plots the self-resonance frequency of stacked and

[1]The skin effect and substrate loss appear to be the limiting factors at these frequencies.

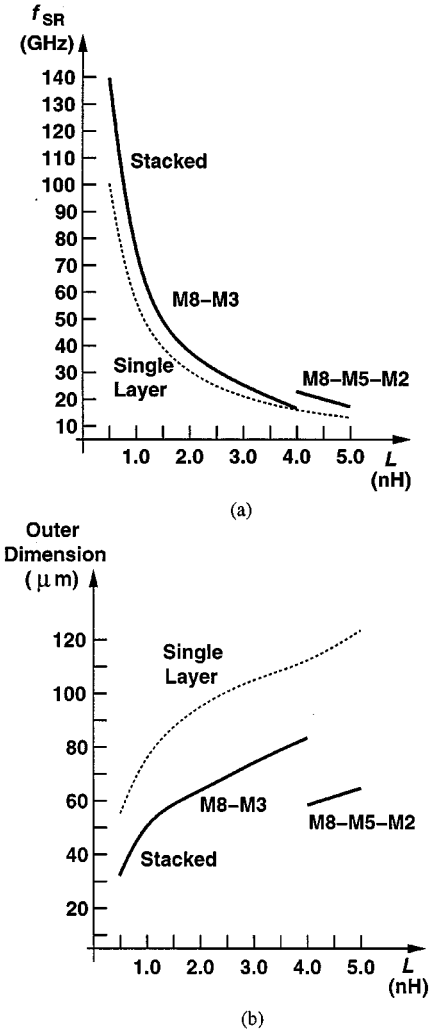

Fig. 5. (a) Self-resonance frequency. (b) Outer dimension of inductors.

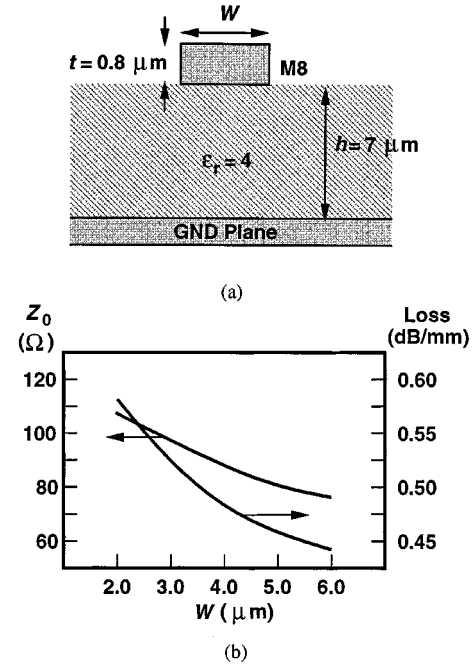

Fig. 6. (a) Microstrip structure in CMOS technology. (b) Loss and characteristic impedance as a function of linewidth.

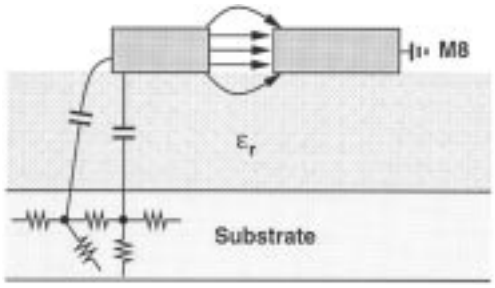

Fig. 7. Coplanar T line.

single-layer inductors ranging from 0.5 to 5 nH. Here, the inductance is obtained by ASITIC simulations and the equivalent capacitance from the derivations in [2]. As described in [2], increasing the vertical spacing between spirals can substantially raise $f_{\rm SR}$ without degrading other properties of the inductor, hence, the use of M8 and M3 rather than M8 and M7. Fig. 5(b) plots the outer dimension of the same inductors, demonstrating the significant area savings provided by stacking. Combinations of parallel and series spirals also prove attractive for increasing the Q.

Techniques such as octagonal spirals [3], patterned shields [4], parallel spirals [5], and tapering the linewidth of the inductor from inner turns to outer turns have been proposed to increase the Q of inductors, but the improvement heavily depends on the frequency of operation. For this reason, fabrication and measurement of many such topologies are often necessary to determine the optimum structure with respect to the Q, area, and $f_{\rm SR}$.

Transmission Lines: CMOS processes can now afford structures that have frequently been used in III–V technologies for microwave and millimeter-wave applications. Specifically, the multitude of metal layers provides transmission lines with a reasonable loss and a small capacitance per unit length, two properties essential to active circuit design using T lines.

Fig. 6(a) shows a microstrip structure consisting of a metal-8 line over a metal-1 ground plane. Plotted in Fig. 6(b) are the loss at 40 GHz and the characteristic impedance versus the linewidth, obtained by electromagnetic field simulations. As a typical example, $W = 2\ \mu$m yields a loss of 0.58 dB/mm and a characteristic impedance of 110 Ω. This performance is acceptable for high-speed distributed amplifiers and oscillators.

Fig. 7 depicts a coplanar T line in CMOS technology [6]. A wide spacing between the signal and ground lines translates to a relatively small capacitance, but the field lines terminating on the substrate may introduce significant loss. Furthermore, such a structure both *creates* and/or *senses* more substrate noise than the microstrip counterpart.

MOS Varactors: Another component inherited from RF CMOS design is the MOS varactor. Shown in Fig. 1, the device is realized as an nFET placed inside an n-well, thereby exhibiting a monotonic C–V characteristic [7], [8]. In contrast to p-n-junction varactors, MOS varactors can sustain both negative and positive voltages, yielding a wider tuning range for VCOs, especially at low supply voltages.

It is unclear whether MOS varactors provide a higher or lower Q than do p-n junctions at tens of gigahertz, but measurements

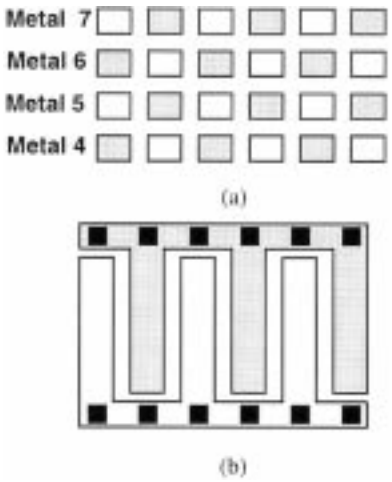

Fig. 8. Fringe capacitor. (a) Side view. (b) Top view.

on a 40-GHz *LC* CMOS VCO indicate that the tank Q is still determined by that of the inductor, suggesting that the MOS varactor's Q does not limit the performance.

Fringe Capacitors: 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. 8 [10], 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 fF/μm can be achieved.

C. Device Modeling

A number of practical issues limit modeling capabilities at high speeds, making fabrication and measurement an essential part of device design.

The principal difficulty in modeling MOSFETs for OC circuits relates to their thermal and flicker noise. The excess noise coefficient γ in $\overline{I_n^2} = 4kT\gamma g_m$ and the flicker noise corner frequency of short-channel transistors must often be obtained by direct device measurements for each technology generation. These noise characteristics play a significant role in the performance of VCOs, transimpedance amplifiers (TIAs), limiting amplifiers, phase detectors, and charge pumps.

Interestingly, the capacitances of MOS devices represented by BSIM3 models appear to be relatively accurate even at tens of gigahertz. This is validated by experimental characterization of a number of *LC* oscillators in the range of 15–40 GHz: Simulated and measured oscillation frequencies differ by less than 5%.

Passive device modeling entails a number of difficulties. Inductor characterization programs using electromagnetic field simulations typically fail to accurately model skin and substrate effects at frequencies above approximately 5 GHz. Also, as shown in Fig. 9, actual T lines in CMOS technology contain multiple dielectric constants and a finite ground plane surrounded by a lossy substrate. Field simulators may not be able to handle such complex geometries.

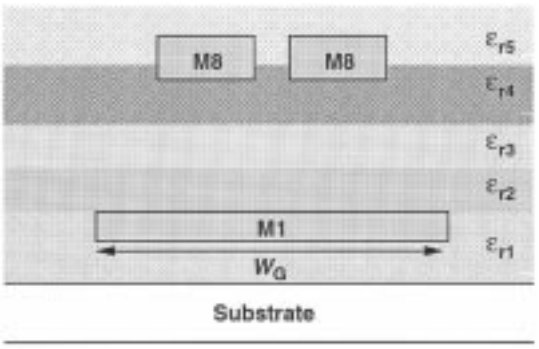

Fig. 9. Actual differential T line in CMOS technology.

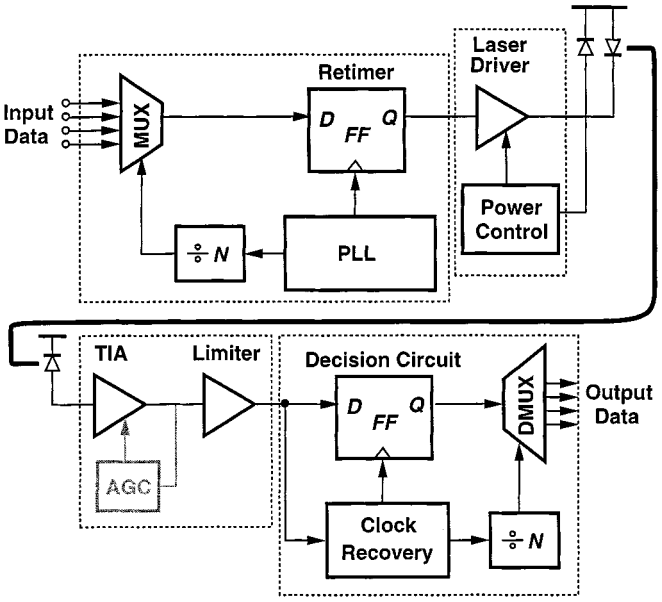

Fig. 10. Traditional OC transceiver.

Simple oscillators can serve as vehicles for inductor (and T line) characterization at high frequencies. As explained in [9], matching the simulated oscillation frequency and device transconductance to the measured value leads to a simple inductor model.

III. Traditional OC Transceivers

Fig. 10 shows a traditional optical system. In the transmitter (TX), a number of channels are multiplexed into a high-speed data stream, the result is retimed and applied to a laser driver, and the optical output is delivered to the fiber. A phase-locked loop (PLL) generates clocks for both the multiplexer and the retiming flip-flop (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 TIA and a limiter[2] raise the signal swing to logical levels. [The TIA may incorporate automatic gain control

[2]RF and optical communication design use the terms "limiter" and "limiting amplifier," respectively, to refer to the same circuit. We use the two interchangeably here.

(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. 10 entails several issues that manifest themselves at high speeds and/or in scaled integrated circuit (IC) technologies. Since the jitter of the transmitted data is determined by primarily that of the PLL, a robust low-noise design 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 (or modulator) driver, a circuit that must deliver tens of milliamperes of current with very short rise and fall times. Since laser diodes or modulators 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 [1]. In contrast to RF power amplifiers, laser drivers must operate across a broad frequency range, prohibiting the use of matching networks with limited bandwidths. For these reasons, high-power laser and modulator drivers may remain outside the realm of deep-submicron CMOS design.

The receiver of Fig. 10 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. 10 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 PLL and the clock and data recovery (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 PLL 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. 10 indicate a typical partitioning, suggesting the following single-chip blocks:

- PLL/MUX circuit (also called the "serializer");
- laser driver along with its power control circuitry;
- TIA/limiter combination;
- CDR/MUX circuit (also called the "deserializer").

Recent work has integrated the serializer and deserializer (producing a "serdes") but the TX and RX front-end amplifiers tend to remain in isolation.

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 with low-voltage differential signaling (LVDS) outputs across 100-Ω differential loads typically draws a supply current of 16×5 mA $= 80$ mA, which is a significant fraction of the overall transceiver's current. Thus, if the supply voltage is decreased from 3 to 1.2 V, the DMUX power dissipation drops considerably.

The second attribute relates to the cost. Owing to the 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 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 toward 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. Section IV provides examples.

IV. Modern OC Transceivers

In response to the demand for higher data rates, the Optical Internetworking Forum (OIF) has proposed two solutions for 40-Gb/s communication networks.[3] The first incorporates wave-division multiplexing (WDM), creating four channels on a single fiber and carrying a data rate of 10 Gb/s in each channel. The second assumes a single wavelength carrying a data rate of 40 Gb/s.

A. Quad 10-Gb/s Transceivers

The OIF proposal for WDM requires a TRX that is *more* than four times as complex as a single 10-Gb/s transceiver. Conceptually illustrated in Fig. 11(a), the TRX employs four serdes with 16 2.5-Gb/s parallel channels on the low-speed side and four 10-Gb/s channels on the high-speed side. The 2.5-Gb/s channels are applied to and received from the framer, which is presently assumed to be on a separate chip.

The communication between the framer and the four serdes at 2.5 Gb/s and over traces on a printed-circuit board (PCB) leads to two critical issues. First, since it is difficult to guarantee equal trace lengths for the 16 channels, the PCB inevitably introduces significant skews. For example, with a propagation velocity of 2.5×10^8 m/s, a 1-cm length difference yields a 40-ps skew. Second, the skews may vary with manufacturing and temperature, mandating continuous correction.

In order to resolve the above issues, OIF has proposed the Serdes-Framer Interface, Level 5 (SFI-5). Used in both the quad TRX and the framer, SFI-5 accommodates relatively long skews and performs continuous skew correction during data communication. Fig. 11(b) depicts the interface. In addition to the 16 data channels, the framer generates and transmits a deskew channel

[3][Online.] Available: http://www.oiforum.com.

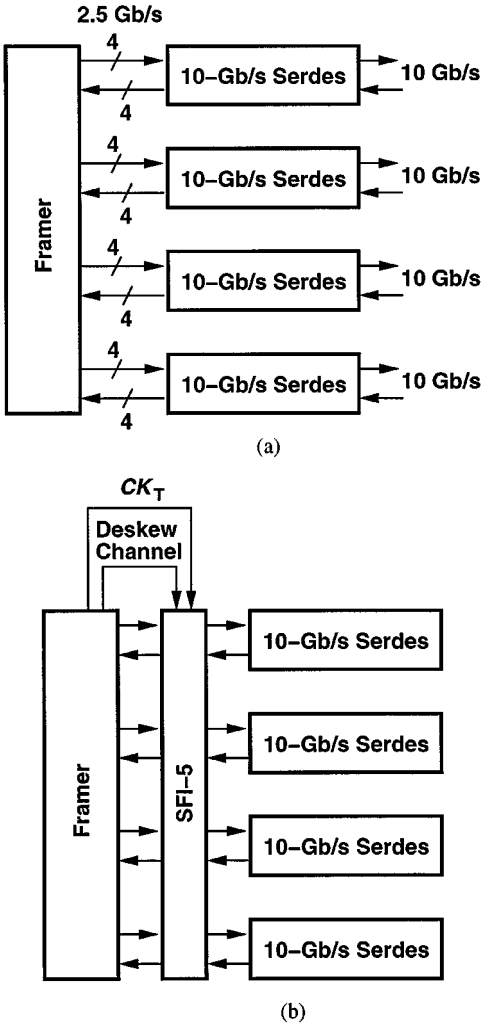

Fig. 11. (a) Interface between a framer and a quad serdes. (b) Addition of SFI-5.

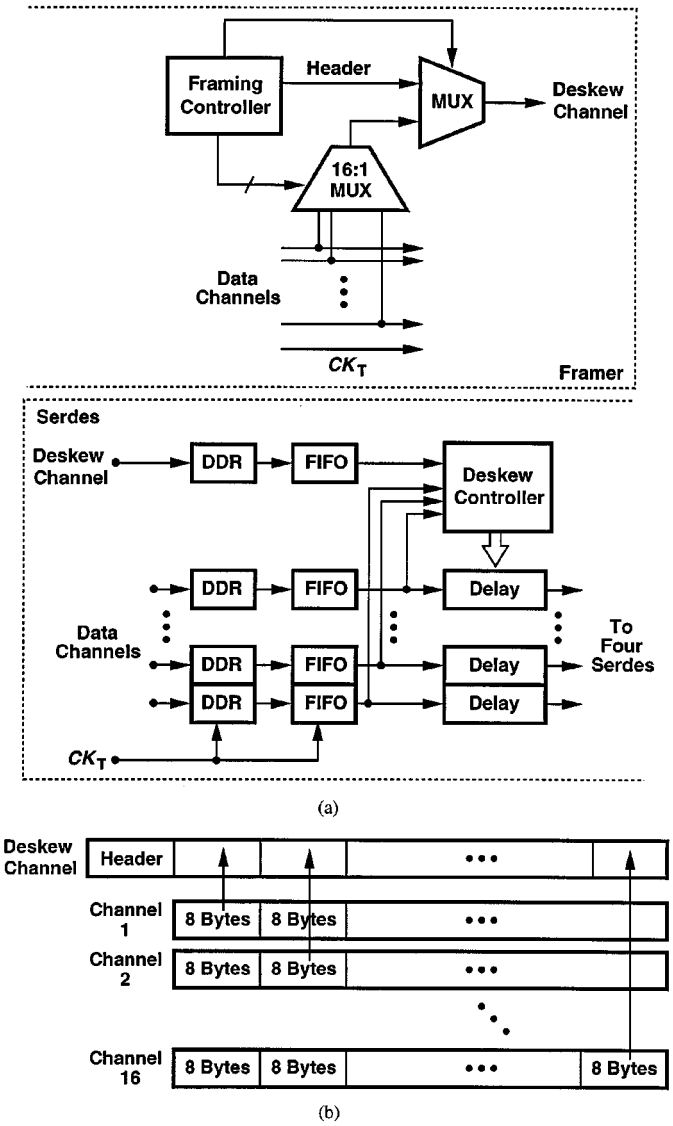

Fig. 12. (a) Details of SFI-5. (b) Deskew frame structure.

and a clock CK_{T}. The serdes then utilizes this channel to insert enough delay (in discrete steps equal to one bit) in each data channel, thus aligning all 16.

A more detailed view of SFI-5 is depicted in Fig. 12(a). To generate the deskew channel, a 16-to-1 MUX in the framer sequentially captures eight bytes from each data channel. A 2-to-1 MUX then allows the header to precede the data bytes. The overall operation is coordinated by the "framing controller," which also generates the header.

On the serdes side, first digital data recovery (DDR) is performed. DDR can employ an analog delay-locked loop (DLL) to align the bits in each channel with the clock, thereby enabling optimum sampling by subsequent stages and avoiding metastability. The retimed channels are then applied to first-in–first-out (FIFO) registers so as to absorb small errors between the clock frequencies on the framer and the serdes.

Following the above operations, the task of deskewing is carried out. Each programmable delay chain consists of a number of flip-flops that can delay the data or be bypassed. The deskew controller continuously monitors the data channels and the deskew channel, adjusting each programmable delay chain to guarantee alignment of data at the input of the serializer.

The real-time skew correction required by SFI-5 leads to a complex frame structure for the deskew channel. Each frame in this channel consists of 1) a header, which indicates the beginning of the frame, and 2) a sample of random data carried by each of the 2.5-Gb/s channels. Fig. 12(b) shows the overall frame. The 8-byte header has a predefined format and is followed by eight bytes from channel 1, eight bytes from channel 2, etc.

SFI-5 is a relatively complex digital machine, requiring roughly 5000–10 000 flip-flops that operate at 2.5 GHz. In addition to limited timing budget for logical operations, SFI-5 poses many other challenges in the design: the large number of flip-flops running at this speed consume substantial power, present a large capacitance to the 2.5-GHz clock, generate considerable supply noise, and inject a great deal of noise into the substrate. Proper choice of the logic style, applying analog layout techniques, and the use of deep n-well can alleviate these issues.

The quad 10-Gb/s TRX with SFI-5 serves as a compelling case for the use of CMOS technology. The complexity of four

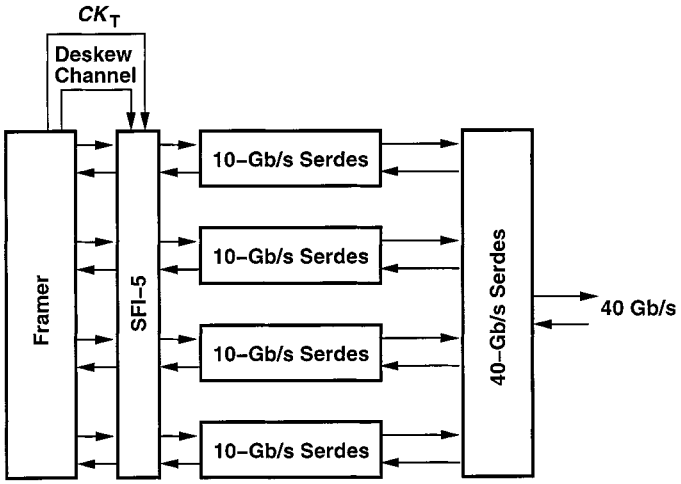

Fig. 13. SFI-5 used in a 40-Gb/s system.

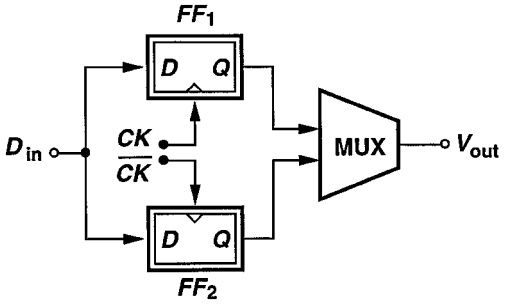

Fig. 14. Multiplexed flip-flops.

serdes and SFI-5 can be accommodated by standard CMOS technologies at a low cost. Furthermore, with a supply voltage of about 1.2 V in 0.13-μm processes, the overall power dissipation can be maintained below 3 W.

B. 40-Gb/s Transceivers

The next generation of SONET, namely, OC768, requires transceivers operating with 40-Gb/s serial data.[4] The 2.5-Gb/s parallel interface may still be based on SFI-5, yielding the architecture shown in Fig. 13. We describe two important issues in this system.

The difficulties in the design of 40-Gb/s circuits make it desirable to perform some of the operations at half rate. For example, the CDR function in the receiver can utilize a 20-GHz VCO along with a half-rate phase detector, thereby recovering the data and demultiplexing it into two channels [11], [12]. Similarly, the transmitter may use a 20-GHz PLL along with two time-interleaved retiming flip-flops whose outputs are multiplexed (Fig. 14). However, the latter remedy can potentially introduce significant jitter in the 40-Gb/s output data. For example, if the 20-GHz clock suffers from duty cycle distortion or if the interleaved flip-flops exhibit mismatches, then the output data incurs pulsewidth distortion. The transmit path must therefore incorporate full-rate circuits.

The second issue in the architecture of Fig. 13 relates to the partitioning of the system according to the capabilities of the IC

[4]The use of coding for forward error correction may raise the rate to 43 Gb/s.

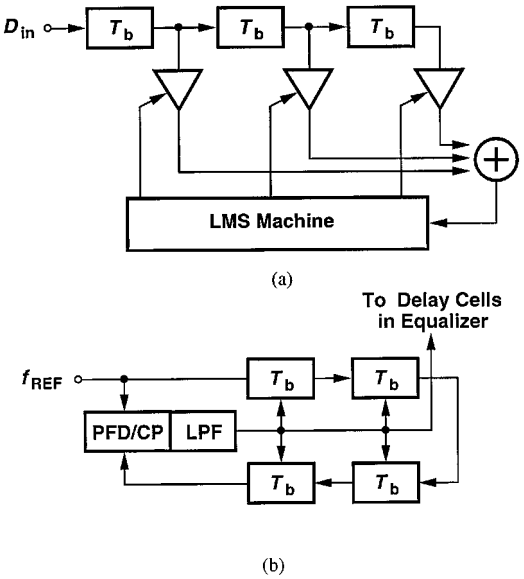

Fig. 15. (a) Tapped delay line equalizer. (b) Use of a DLL to define unit delay.

technologies. Specifically, one may envision partitioning the system into a CMOS quad TRX with SFI-5 [similar to Fig. 11(b)] and a bipolar front-end consisting of a 4-to-1 MUX, a 40-GHz PLL, a 40-Gb/s CDR, and a 1-to-4 DMUX. However, such an arrangement would demand that four (differential) 10-Gb/s channels travel between the two chips while experiencing minimal skew with respect to the clocks on both chips. This problem is similar to the skew issues that have led to the invention of SFI-5, but much more difficult because of the higher data rate. With the physical dimensions of traces and packages, it would be impractical to ensure adequate data and clock alignment between the two chips (unless the bipolar front-end employs a 10-Gb/s version of SFI-5).

The above two issues indicate that, if CMOS technology can support 40-Gb/s data rates, then a single-chip solution having a low cost and low power dissipation becomes feasible. Methods of approaching such speeds in amplifiers and oscillators are presented in Section VI.

V. Equalization

Dispersion in optical fibers has become a serious issue in recent years. Most silica fibers deployed in the 1980s were designated to operate at a wavelength of 1.33 μm, where (material) dispersion drops to zero. Since then, however, most of optical communication has shifted to a wavelength of 1.55 μm for two reasons. First, erbium-doped fiber amplifiers (EDFAs), whose gain window appears around 1.55 μm, have found widespread usage in long-haul applications. Second, the loss of silica fibers reaches a minimum at this wavelength. As a result, dispersion is significant even for a data rate of 10 Gb/s.

For long-haul and/or 40-Gb/s applications, dispersion compensation becomes essential. Two types of dispersion manifest themselves more prominently: 1) chromatic (material) dispersion, which results from the dependence of the refraction index, and hence, the propagation velocity upon the wavelength; and 2) polarization-mode dispersion, which arises from different propagation velocities for different modes of polarization, an effect

due to the deviation of the fiber cross section from a perfect circle (birefringence) [13].

Most of the dispersion compensation research has thus far appeared in the optical domain but yields expensive and bulky solutions. It is, therefore, desirable to exploit the vast knowledge of equalization in signal processing to suppress dispersion in the electrical domain. To remove the intersymbol interference (ISI) due to dispersion (and due to other nonidealities), an adaptive equalization path and an adaptation machine must be interposed between the TIA and the limiting amplifier in Fig. 10. For example, as shown in Fig. 15(a), a tapped delay line along with a least-mean-square (LMS) algorithm can serve to suppress part of the dispersion [14]. At high speeds, however, a number of issues arise here. First, it becomes impractical to employ discrete-time processing of data because no clock has been recovered yet, and the limited timing budget prohibits the use of switching operations. Second, with continuous-time processing, it is difficult to guarantee that each cell in the delay line (e.g., a differential pair) provides a delay equal to one bit period across process and temperature extremes. Thus, it may be necessary to create a replica cell whose delay is defined by a DLL [Fig. 15(b)].

The complexity of the equalizer, especially if it appears four times in a quad 10-Gb/s system, also strengthens the reason for integration in CMOS technology. Interestingly, the use of equalization also relaxes the TIA design. The bandwidth of TIAs is typically chosen to be around 0.7 times the bit rate as a compromise between the total noise and acceptable ISI. With equalization, on the other hand, the TIA bandwidth may be as small as a quarter of the data rate, thereby improving the sensitivity and even enabling a greater transimpedance gain. Nonetheless, the TIA must exhibit some linearity so as to allow subsequent equalization.

VI. High-Speed Techniques

The two principal issues in high-speed CMOS design for OC circuits are the limited f_T and the low supply voltage. In fact, the latter prohibits the use of many well-known techniques such as the Cherry–Hooper topology [15] or the Gilbert gain cell [16].

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 prediction of the performance in optical communication circuits as well. Interestingly, inductor Qs as low as 3–4 prove adequate for increasing the bandwidth, allowing the use of simple, compact spiral structures.

Fig. 16(a) shows a gain 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%, which is still quite a significant factor.

An interesting difficulty in modeling the inductors in the above circuit arises from the narrow-band nature of the definition of the Q, an issue rarely encountered in RF design.

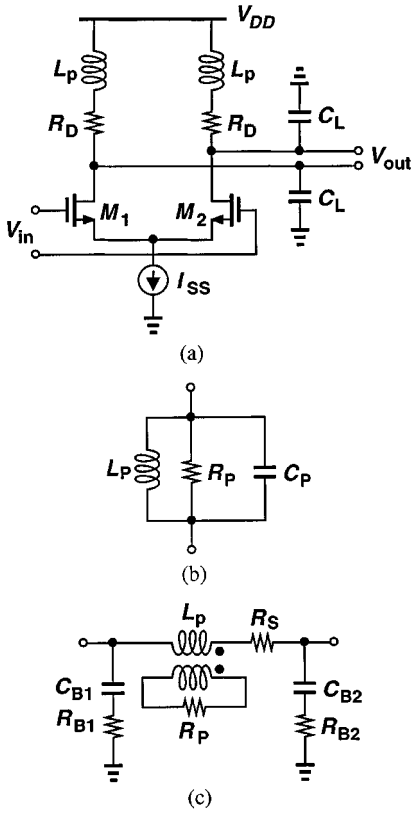

Fig. 16. (a) Inductive peaking. (b) Simple inductor model. (c) More complete inductor model.

Fig. 16(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. 16(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. 16(b) can.

The problem of broad-band amplification becomes much more difficult if the circuit must deliver large currents to off-chip loads because the wide transistors necessary for this task introduce a large input capacitance. This issue can be ameliorated through the use of f_T doublers [17]. Consider the circuit shown in Fig. 17(a), where the device dimensions and bias currents are chosen according to gain and voltage headroom requirements. We wish to modify the circuit such that the input capacitance decreases while the voltage gain remains unchanged. The small-signal behavior of the circuit is expressed as $V_{\rm out} = g_m(V_{\rm in1} - V_{\rm in2})R_D$, where g_m denotes the transconductance of each transistor. Now suppose two such differential pairs are configured as shown in Fig. 17(b), where the input ports are placed in *series* while the output ports are connected in parallel. (The load resistors are still equal to R_D.) The bias voltage V_b is chosen equal to the common-mode level

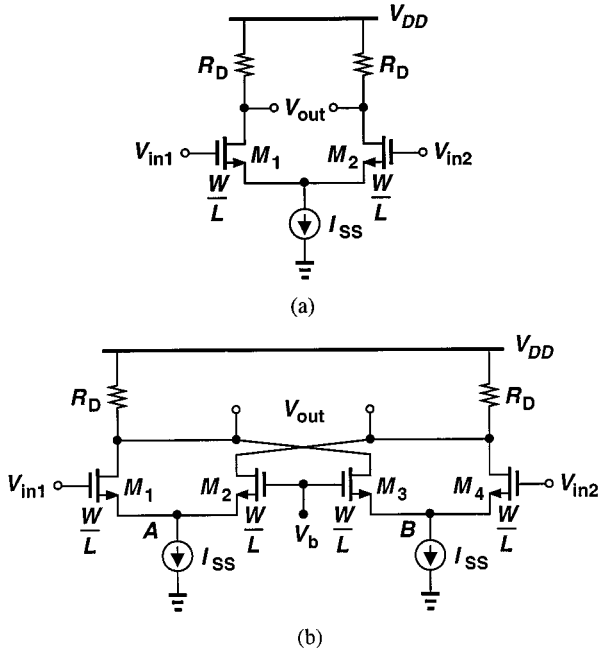

Fig. 17. (a) Simple differential stage. (b) f_T doubler.

of $V_{\rm in1}$ and $V_{\rm in2}$, allowing the differential pairs to operate with zero systematic offset. Using superposition to calculate $V_{\rm out}$ in terms of $V_{\rm in1}$ and $V_{\rm in2}$, we have $V_{\rm out} = g_m(V_{\rm in1} - V_{\rm in2})R_D$. The circuit thus provides the same voltage gain but with a lower input capacitance. In fact, if the parasitic capacitance at nodes A and B is negligible, then the input capacitance seen by $V_{\rm in1}$ or $V_{\rm in2}$ is roughly equal to $C_{\rm GS}/2$. Since this circuit halves the input capacitance while maintaining the same overall transconductance, it is called an f_T doubler.

The f_T doubler of Fig. 17(b) nonetheless suffers from several drawbacks. First, the power dissipation is doubled. Second, the total current flowing through the load resistors is doubled, possibly driving the transistors into the triode region. Third, the total capacitance contributed by the transistors to the *output* nodes is doubled, lowering the output pole. Fourth, if the source–bulk junction capacitance of the transistors and the capacitance introduced by the tail current sources is not negligible, the input capacitance is higher than $C_{\rm GD}/2$.

Despite the above issues, f_T doublers prove useful in broad-band output buffers. Fig. 18(a) shows an example employing high-speed techniques to deliver a differential voltage swing of approximately 340 mV to 75-Ω on-chip termination resistors and 50-Ω off-chip loads. The output stage also utilizes inductive peaking to achieve faster transitions [1]. The simulated eye diagram for a data rate of 40 Gb/s in 0.13-μm technology is shown in Fig. 18(b). The circuit consumes 27 mW from a 1.2-V supply.

B. Oscillators

The speed and noise performance of ring oscillators makes them a poor choice for OC applications. With inductor Qs exceeding 10 at several tens of gigahertz, *LC* VCOs continue to play a critical role in high-speed PLLs and CDR circuits. Fig. 19(a) depicts a VCO incorporating spiral inductors and

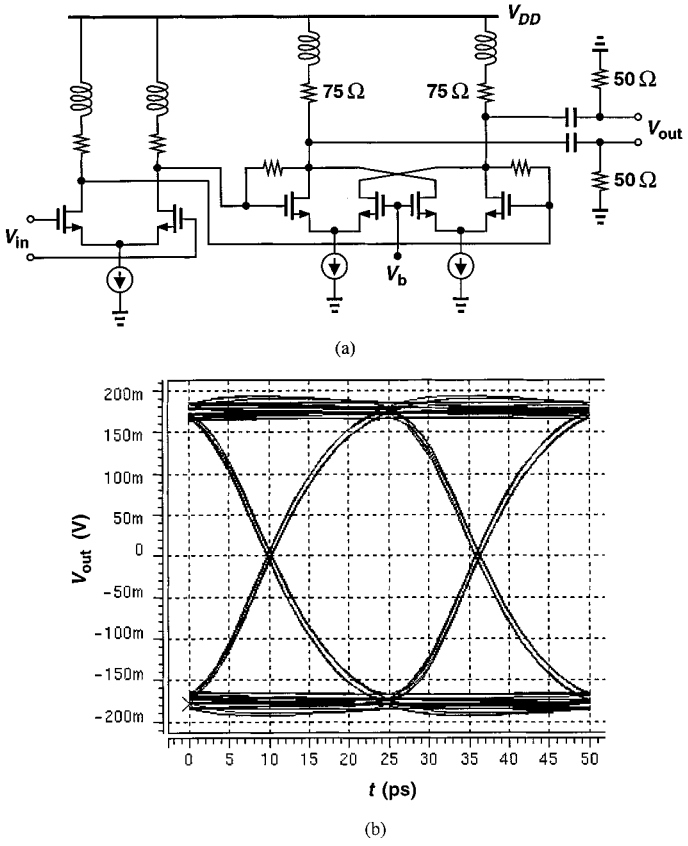

Fig. 18. (a) High-speed output buffer. (b) Simulated output eye diagram.

MOS varactors. Capacitors C_1 and C_2 isolate the dc level applied to the varactors from the output common-mode level. The voltage V_b is approximately equal to $V_{DD}/2$, allowing both positive and negative voltages across the varactors, and hence, maximizing the tuning range. With $L_1 = L_2 = 0.25$ nH, $(W/L)_{1,2} = 4\ \mu$m$/0.13\ \mu$m, and $I_{SS} = 2$ mA, the circuit operates at 40 GHz, providing a differential output swing of 2 V and a tuning range of 4 GHz.

Fig. 19(b) plots the simulated phase noise of the oscillator for $I_{SS} = 1$ and 2 mA. The phase noise at 1-MHz offset is equal to -94 and -98 dBc/Hz, respectively. These values, however, are obtained for an nMOS noise coefficient of 2/3. Since for short-channel devices, γ may reach 2.5, the actual phase noise is higher by approximately $10\log[2.5/(2/3)] = 5.74$ dB.

An important issue in the VCO of Fig. 19(a) relates to the parasitic bottom-plate capacitance of C_1 and C_2 to the substrate. Capacitors realized as a sandwich of metal layers typically suffer from a bottom-plate parasitic of more than 10%. Since C_1 and C_2 must be at least five times the maximum value of the varactors so as to appear "transparent," their parasitics become comparable with the varactor capacitances, thus limiting the tuning range. An attractive candidate for this purpose is the fringe capacitor of Fig. 8 as its bottom-plate capacitance typically falls around a few percent.

The availability of high-performance oscillators in CMOS technology also eases the design of frequency dividers. Recent work on injection-locked oscillators indicates that they can perform frequency division with low power and low noise [18].

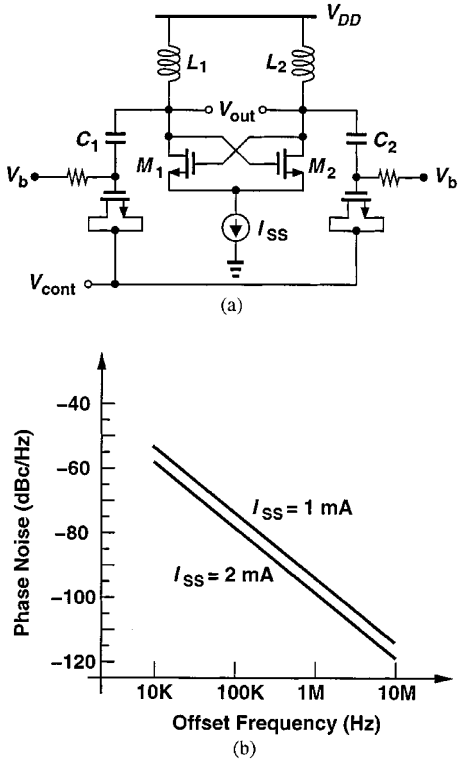

Fig. 19. (a) The 40-GHz *LC* VCO. (b) Phase noise as a function of frequency offset.

C. Relationship Between Phase Noise and Jitter

With the above VCO example, we must now answer the important question of how the jitter of a phase-locked oscillator is related to its free-running phase noise. This question plays a central role in PLL and CDR design for two reasons: 1) the phase noise of oscillators can be simulated and measured much more easily than the corresponding jitter can; and 2) with the relationship known, it is possible to determine the maximum tolerable oscillator phase noise and design the circuit accordingly.

If only the phase noise due to white noise sources is considered, then it can be shown that the rms cycle-to-cycle jitter of a free-running oscillator is related to its phase noise as

$$\Delta T_{cc}^2 \approx \frac{4\pi}{\omega_0^3} S_\phi(\Delta\omega)\Delta\omega^2 \tag{1}$$

where $\omega_0 = 2\pi f_0$ denotes the oscillation frequency and $S_\phi(\Delta\omega)$ represents the relative single-sideband phase noise power at an offset frequency of $\Delta\omega$ [19].

To obtain the jitter of the phase-locked oscillator, it can be assumed that, for a loop bandwidth of $2\pi f_u$, the jitter rises with the square root of time (as if the oscillator were free-running) until $t_1 = (2\pi f_u)^{-1}$ and "saturates" thereafter (Fig. 20) [20]. As proved in [19], the total jitter accumulated over time t_1 by a free-running oscillator is equal to

$$\Delta T_1 = \sqrt{\frac{f_0}{2}}\,\Delta T_{cc}\sqrt{t_1}. \tag{2}$$

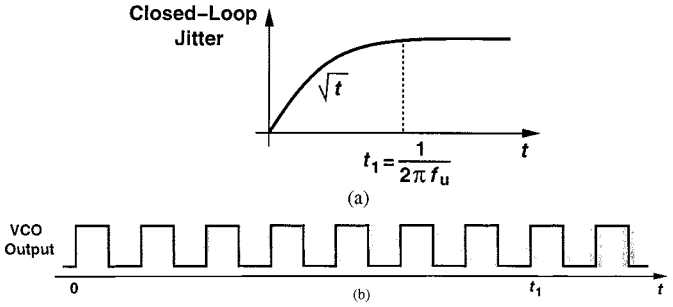

Fig. 20. Accumulation of cycle-to-cycle jitter in a phase-locked oscillator.

Substituting (1) in (2) yields the closed-loop jitter as

$$\Delta T_{\mathrm{PLL}} = \frac{1}{\sqrt{2\pi f_u}} \sqrt{S_\phi(\Delta\omega)}\,\frac{\Delta\omega}{\omega_0}. \tag{3}$$

The above derivations have been verified by behavioral simulations of phase noise and jitter. For example, if ΔT_{PLL} must be less than 0.25 ps, rms, at 40 GHz and $f_u = 20$ MHz, then $S_\phi(\Delta\omega)$ must not exceed -79 dBc/Hz at 1-MHz offset. This indicates that the oscillator example of Fig. 19(a) may provide a reasonable jitter performance at 40 GHz.

VII. Conclusion

CMOS technology offers many advantages for the integration of modern OC transceivers. The multitude of metal layers, the deep n-well, and the MOS varactor structure prove invaluable in extending the capabilities of basic CMOS devices to both higher speeds and greater levels of integration. Moreover, the low supply voltage translates to a lower power dissipation for most of the transceiver building blocks. With the growing port density in OC systems, these features of CMOS technology can lead to low-cost efficient solutions.

References

[1] H.-M. Rein and M. Moller, "Design considerations for very high-speed Si bipolar ICs operating up to 50 Gb/s," *IEEE J. Solid-State Circuits*, vol. 31, pp. 1076–1090, Aug. 1996.

[2] A. Zolfaghari, A. Y. Chan, and B. Razavi, "Stacked inductors and transformers in CMOS technology," *IEEE J. Solid-State Circuits*, vol. 36, pp. 620–628, Apr. 2001.

[3] B. De Muer, M. Borremans, M. Steyaert, and G. Li Puma, "A 2-GHz low phase noise integrated *LC* VCO set with flicker noise upconversion minimization," *IEEE J. Solid-State Circuits*, vol. 35, pp. 1034–1038, July 2000.

[4] C. P. Yue and S. S. Wong, "On-chip spiral inductors with patterned ground shields for Si-based RF ICs," *IEEE J. Solid-State Circuits*, vol. 33, pp. 743–752, May 1998.

[5] R. B. Merril, T. W. Lee, H. You, R. Rasmussen, and L. A. Moberly, "Optimization of high-Q inductors for multilevel metal CMOS," in *Proc. IEDM*, Dec. 1995, pp. 38.7.1–38.7.4.

[6] B. Kleveland, C. H. Diaz, D. Vock, L. Madden, T. H. Lee, and S. S. Wong, "Monolithic CMOS distributed amplifier and oscillator," in *IEEE Int. Solid State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 1999, pp. 70–71.

[7] C.-M. Hung, Y.-C. Ho, I.-C. Wu, and K. O, "High-Q capacitors implemented in a CMOS process for low-power wireless applications," *IEEE Trans. Microwave Theory Tech.*, vol. 46, pp. 505–511, May 1998.

[8] A. S. Porret, T. Melly, C. C. Enz, and E. A. Vittoz, "Design of high-Q varactors for low-power wireless applications using a standard CMOS process," *IEEE J. Solid-State Circuits*, vol. 35, pp. 337–345, Mar. 2000.

[9] B. Razavi, "CMOS technology characterization for analog and RF design," *IEEE J. Solid-State Circuits*, vol. 34, pp. 268–276, Mar. 1999.
[10] O. E. Akcasu, "High-capacity structures in a semiconductor device," U.S. Patent 5 208 725, May 4, 1993.
[11] M. Wurzer, J. Böck, H. Knapp, W. Zirwas, F. Schumann, and A. Felder, "A 40-Gb/s integrated clock and data recovery circuit in a 50-GHz f_T silicon bipolar technology," *IEEE J. Solid-State Circuits*, vol. 34, pp. 1320–1324, Sept. 1999.
[12] J. Savoj and B. Razavi, "A 10-Gb/s CMOS clock and data recovery circuit with frequency detection," in *IEEE Int. Solid State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2001, pp. 78–79.
[13] I. P. Kaminow and T. L. Koch, Eds., *Optical Fiber Telecommunications IIIA*. San Diego, CA: Academic, 1997.
[14] J. H. Winters and R. D. Gitlin, "Electrical signal processing techniques in long-haul fiber-optic systems," *IEEE Trans. Commun.*, vol. 38, pp. 1439–1453, Sept. 1990.
[15] E. M. Cherry and D. E. Hooper, "The design of wide-band transistor feedback amplifiers," *Proc. Inst. Electr. Eng.*, vol. 110, pp. 375–389, Feb. 1963.
[16] B. Gilbert, "A new wideband amplifier technique," *IEEE J. Solid-State Circuits*, vol. SSC-3, pp. 353–365, Dec. 1968.
[17] D. Feucht, *Handbook of Analog Circuit Design*. San Diego, CA: Academic, 1990.
[18] H. Rategh and T. H. Lee, "Superharmonic injection-locked frequency dividers," *IEEE J. Solid-State Circuits*, vol. 34, pp. 813–821, June 1999.
[19] F. Herzel and B. Razavi, "A study of oscillator jitter due to supply and substrate noise," *IEEE Trans. Circuits Syst. II*, vol. 46, pp. 56–62, Jan. 1999.
[20] J. A. McNeill, "Jitter in ring oscillators," *IEEE J. Solid-State Circuits*, vol. 32, pp. 870–879, June 1997.

Behzad Razavi (S'87–M'90–SM'00) 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 in 1995. He was with AT&T Bell Laboratories and Hewlett-Packard Laboratories until 1996. Since September 1996, he has been an 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* (New York: IEEE Press, 1995), *RF Microelectronics* (Englewood Cliffs, NJ: Prentice-Hall, 1998), *Design of Analog Integrated Circuits* (New York: McGraw-Hill, 2001), *Design of Integrated Circuits for Optical Communications* (New York: McGraw-Hill, 2002), and the editor of *Monolithic Phase-Locked Loops and Clock Recovery Circuits* (New York: IEEE Press, 1996). 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 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 corecipient of the Jack Kilby Outstanding Student Paper Award at the 2002 ISSCC. He served on the Technical Program Committee of the International Solid-State Circuits Conference (ISSCC) from 1993 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*.