

On-Chip Oscilloscopes for Noninvasive Time-Domain Measurement of Waveforms in Digital Integrated Circuits

Yu Zheng, *Student Member, IEEE*, and Kenneth L. Shepard, *Member, IEEE*

Abstract—High-speed digital design is becoming increasingly analog. In particular, interconnect response at high frequencies can be nonmonotonic with “porch steps” and ringing. Crosstalk (both capacitive and inductive) can result in glitches on wires that can produce functional failures in receiving circuits. Most of these important effects are not addressed with traditional automatic test pattern generation (ATPG) and built-in self-test (BIST) techniques, which are limited to the binary abstraction. In this work, we explore the feasibility of integrating primitive sampling oscilloscopes on-chip to provide waveforms on selective critical nets for test and diagnosis. The oscilloscopes rely on subsampling techniques to achieve 10-ps timing accuracy. High-speed samplers are combined with delay-locked loops (DLLs) and a simple 8-bit analog-to-digital converter (ADC) to convert the waveforms into digital data that can be incorporated as part of the chip scan chain. We will describe the design and measurement of a chip we have fabricated to incorporate these oscilloscopes with a high-frequency interconnect structure in a TSMC 0.25- μm process. The layout was extracted using Cadence’s Assura RCX-PL extraction engine, enabling a comparison between simulated and measured results.

Index Terms—Inductance modeling, mixed-signal test, subsampling.

I. INTRODUCTION

THERE is strong recent interest in the ability to noninvasively measure waveforms in the time-domain in integrated circuits. In digital design, this interest stems from the inability of traditional digital test methodologies [e.g., automatic test pattern generation (ATPG) and built-in self test (BIST)] to address the more analog issues of high-speed design such as crosstalk noise and complex nonmonotonic waveforms resulting from the inductive response of high-speed interconnect. E-beam probing and picoprobeing are the only alternatives commonly available for measuring analog waveforms; these techniques are expensive, difficult due to the need to have top-level metal available for probing, and frequently invasive. Moreover, the advent of systems-on-a-chip design is driving the need for testing analog blocks embedded within largely digital integrated circuits [1].

Manuscript received February 27, 2002; revised October 10, 2002. This work was supported in part by the MARCO/DARPA C2S2 Center under subcontract from Carnegie-Mellon University, in part by the National Science Foundation under Grant CCR-97-34216, in part by a gift from the IBM Corporation, and in part by a gift from Cadence Design Systems.

The authors are with the Columbia Integrated Systems Laboratories, Department of Electrical Engineering, Columbia University, New York, NY 10027 USA.

Digital Object Identifier 10.1109/TVLSI.2003.812313

Our motivation for this work has been focused on characterizing the response of on-chip interconnect to provide validation for recently-developed interconnect extraction tools [2]. There has already been considerable work looking at characterizing on-chip wires. Much of this work has been focused on frequency-domain S-parameter characterization [3], [4] generally performed with a high-frequency network analyzer and ground-signal-ground (GSG) probes. While an open pad calibration structure on chip is generally adequate to de-embed the pad parasitics, there are several disadvantages to this approach. First, special pads must be made available for probing and the interconnect structure can not be embedded in a circuit environment with drivers and receivers. Second, the time-domain behavior must be inferred from simulation, in which the inverse fast Fourier transform (IFFT) is used to form the convolution integral for conversion to the time domain.

To avoid the second limitation, direct time-domain measurements have also been made. Deutsch [5] measures the time-domain response to a step excitation directly through high-frequency probes and a sampling oscilloscope. Line delay is extracted by subtracting out the delay of a short reference line, thereby (theoretically) eliminating the effects of the probes, cables, and pads. Crosstalk is observed directly without correction. Unfortunately, this approach still has the disadvantage of requiring special structures for probing and cannot be used to measure “real” wires embedded in a complex circuit environment. To bring the sampling on-chip to allow the measurement to be done on wires embedded in a true circuit environment, Soumyanath [6] uses on-chip comparators and relies on the comparator switch point to determine the sample time. The difficulty is that this time must be calibrated through an off-chip delay path, resulting in a complex external measurement setup with limited timing resolution.

Other previous work, primarily in the context of mixed signal test, has also considered employing on-chip samplers and on-chip samplers with A/D conversion [7]–[11]. These approaches, however, have also relied on external clocks to generate the sample clocks, limiting achievable timing resolution and making for a complex off-chip measurement environment. In this work, we combine high-bandwidth samplers and on-chip A/D conversion with a digital-to-time converter to produce the first fully-integrated digital oscilloscope on-chip [12]. In Section II, we review subsampling as the key to measuring fast waveforms on-chip. Sections III–V consider the key components of our oscilloscope—the samplers, the digital-to-time converter, and the analog-to-digital converter

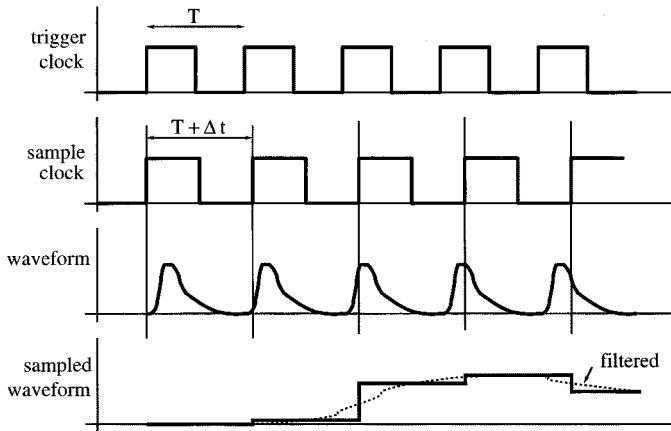


Fig. 1. Subsampling in the time domain. The waveform is spread-out by a factor of $T/\Delta T$.

(ADC), respectively. Section VI presents the overall test chip design. Measurement results on interconnect structures probed with our on-chip oscilloscopes are presented in Section VII. Conclusions are presented in Section VIII.

II. SUBSAMPLING TECHNIQUES

Deep submicron MOS transistors today have f_T 's beyond 50 GHz, making it possible to generate very high bandwidth signals on-chip. One can imagine two approaches to carrying the information in a high-bandwidth signal off-chip. With a fixed (but presumably known) latency, one could buffer the digital data off-chip, but all of the analog information would be lost. To preserve the shape of the waveform, one could use an amplifier with unity-gain feedback to buffer the signal off-chip, but practical bandwidth limitations of the amplifier would limit the signal frequencies that could be sensed to hundreds of megahertz. Fundamentally, the challenge of on-chip measurement circuits is that the circuits performing the measurement are in the same technology as the circuits being measured and, therefore, cannot be made intrinsically "faster."

The key to being able to measure fast waveforms is subsampling. This approach is used in digital sampling oscilloscopes and has been employed in several contexts previously for on-chip measurement circuits [6]–[11]. The approach can be understood from both a time-domain and frequency-domain perspective. From a time-domain point-of-view, imagine that we have two clocks, one of period T (and frequency $f_0 = 1/T$), which we call the *trigger clock* and the other of period $T + \Delta t$ (and frequency $f_s = 1/(T + \Delta t)$), which we call the *sample clock*. We assume that the waveform that we wish to measure is triggered by the leading edge of the signal clock, as shown in Fig. 1, and as such is repeated once each T seconds. If we assume that the sample clock samples the data on its leading edge (and that the sample-and-hold circuit holds the sampled value), then a new time point is sampled each time the waveform is repeated. The output of the sample-and-hold circuit is, therefore, a "spread-out" version of the waveform we wish to measure (as shown in Fig. 1), allowing the ADC or other circuits processing the data to be very low bandwidth.¹

¹In fact, the time scale is magnified by a factor of $T/\Delta t$.

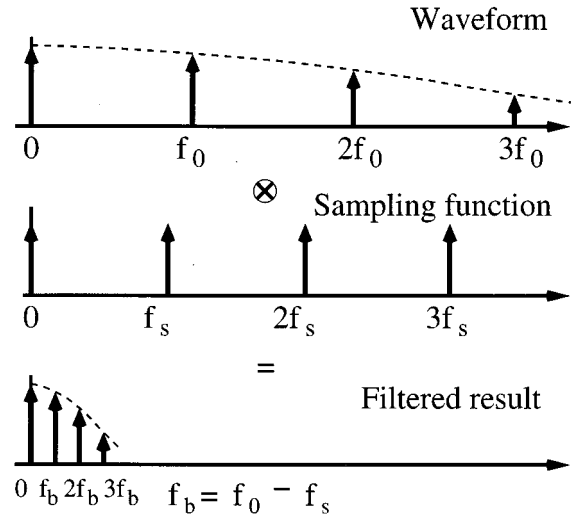


Fig. 2. Subsampling in the frequency domain. The waveform is mixed to multiples of the beat frequency $f_b = f_0 - f_s$.

From a frequency-domain perspective, the waveform to be sampled has a discrete frequency spectrum at multiples of f_0 as shown in Fig. 2(a). The spectrum is discrete because it is periodic with the trigger clock. The sampling process is tantamount to mixing with the frequency spectrum of the sampling function, also discrete at multiples of f_s [Fig. 2(b)]. For an "ideal" sampler with infinite bandwidth, this magnitude is unity at each f_s multiple. The result of this mixing process is a downshifted spectrum at multiples of the beat frequency $f_b = f_0 - f_s \cdot f_s$ is clearly chosen close to f_0 , so that $f_b/f_0 \cong \Delta t/T$ is small. A low-pass filter removes the frequency content above f_s . The action of this filter is shown in the time domain in Fig. 1. Shifted to lower frequency, the signal is easier to measure with "slower" circuits.

On first glance, it would appear that subsampling approaches only apply to "periodic" signals. However, any signal that can be repeatedly generated from a clock edge can be rendered periodic and is amenable to subsampling techniques.

III. SAMPLERS

A critical circuit to the subsampling technique is the sample-and-hold unit. This is the only circuit component of the on-chip oscilloscope that must have a high bandwidth (small aperture window) since it must be able to quickly capture the voltage at the sample clock edge. For this work, we consider samplers based on a master-slave configuration, similar to a master-slave flip-flop. One possible sampler circuit is shown in Fig. 3 [7]. The "master" consists of a nFET pass transistor $M1$ feeding a pFET source-follower unity-gain amplifier. The "slave" is a full-pass transistor feeding a second pFET source-follower. The pFET source follower stages provide several advantages. In addition to nearly unity gain, the use of pFET transistors limit the effect of substrate noise. Also the output range of the buffer matches nicely the input range of an nFET differential pair in the preamplifier stage of a comparator. For fast sample clock transitions, the bandwidth of the sampler is dominated by the time constant of transistor $M1$ charging the capacitance of node $N1$. Transistor $M2$ (at half

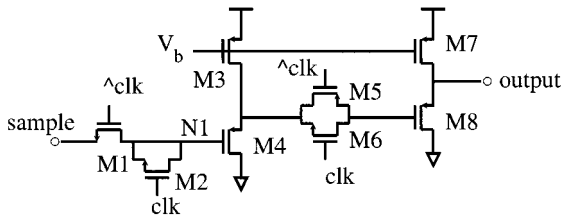


Fig. 3. Master-slave sample-and-hold with pFET source-follower buffers on both master and slave.

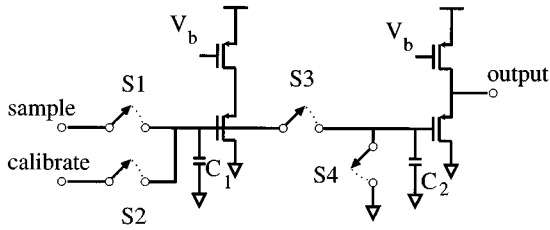


Fig. 4. Charge-sharing, master-slave sample-and-hold.

the width of $M1$ is present to help cancel clock-feedthrough and charge-injection noise associated with $M1$. The main limitation of this sampler is that the source-follower buffers cutoff at input voltage greater than $V_{DD} - |V_{Tp}|$ and, therefore, one cannot sample full-rail signals.²

The range limitation of the sampler of Fig. 3 can be avoided if the buffer is removed from the master, as shown in Fig. 4. This is a variation of the sampler used in [10] and is the sampler used in our testchip. Each of the switches is implemented as a full-pass transistor to ensure high linearity and limit any frequency-dependent distortion.³ In this sampler, charge-sharing between the implicit capacitances C_1 and C_2 divides down the input voltage to be below the $V_{DD} - |V_{Tp}|$ cutoff of the unity-gain buffer.⁴ The usable input range of this sampling “head” is from approximately -300 mV to $V_{DD} + 300$ mV. The samplers are very small, consuming only $100 \mu\text{m}^2$.

The sampler is calibrated with a separate calibrate input, driven by an off-chip reference. This allows calibration of the entire measurement path to digital output, eliminating errors due to analog mismatch, nonlinearities, and offset in both the sampler and the ADC. In sample mode, S_2 is off and the sampler samples the voltage through S_1 . In calibration mode, S_1 is off and a dc voltage is applied off-chip for calibration through S_2 . Because of this calibration, precise matching of C_1 and C_2 in Fig. 4 is not necessary and very small devices can be used. A small C_1 (in our case 30 fF including the parasitic source-drain capacitance of the switches) maximizes the sampler bandwidth and keeps the measurement as noninvasive as possible.

The “master” sampler defined by S_1 and C_1 in Fig. 4 determines the effective bandwidth of the sampling head of our

²Of course, the use of a single nFET switch for the master limits the maximum value of the sampled input to $V_{DD} - V_{Tn}$, but the buffer limitation would remain even if a complementary pass-transistor switch were used.

³Note that the first source follower is a “dummy” to match the capacitive loading of the master to the slave. Capacitors C_1 and C_2 are implicitly created by the device loading.

⁴A variation of this charge-sharing master-slave approach is used in [11]. In this case, the sampled voltage is converted to a current and amplified by a current mirror to be driven off-chip for measurement.

on-chip oscilloscope. The sampler performs a “weighted” average of the sampled signal across the aperture window, determined by both the tracking speed of the switch and the transition time from track to hold. Following the analysis of Johansson [13], the sampled voltage is given by

$$v_{\text{sample}}(t_s) = \int_{-\infty}^{\infty} v_{\text{in}}(\tau)h(\tau - t_s)d\tau \quad (1)$$

where $h(t)$ is the *sampling function*; t_s is the sample time, chosen as a fixed time-reference point in the gate waveform (e. g., the start of its falling transition); and $v_{\text{in}}(t)$ is the sampled voltage waveform. If $v_{\text{in}}(t)$ in simulation is chosen as a small ideal step of magnitude V_o , then

$$v_{\text{sample}}(t_s) = V_o \int_{-\infty}^{\infty} u(\tau)h(\tau - t_s)d\tau \quad (2)$$

Differentiating with respect to t_s yields

$$\begin{aligned} \frac{dv_{\text{sample}}}{dt_s}(t_s) &= V_o \frac{d}{dt_s} \int_{-\infty}^{\infty} u(\tau)h(\tau - t_s)d\tau \\ &= V_o \frac{d}{dt_s} \int_{-t_s}^{\infty} h(\tau)d\tau = V_o h(-t_s) \end{aligned} \quad (3)$$

therefore,

$$h(t_s) = \frac{1}{V_o} \frac{d}{dt_s} v_{\text{sample}}(-t_s). \quad (4)$$

Fig. 5(a) shows $v_{\text{sample}}(t_s)$ from circuit simulation with the actual sample clock drive path included in the simulation.⁵ The corresponding sampling function is given in Fig. 5(b) and its Fourier spectrum is given in Fig. 5(c). Defining the aperture time as the width of the peak of the impulse response in which 80% of the sensitivity is confined yields an aperture time of approximately 60 ps. The -3 -dB bandwidth is approximately 4 GHz. For reference, this technology has an nFET f_T of approximately 30 GHz and a fan-out-of-four (FO4) delay of approximately 100 ps. In Section VII, we will consider the (small) effect the sample bandwidth has on the actual measured results.

IV. DIGITAL-TO-TIME CONVERTER

One of the limitations of the subsampling approach illustrated in Fig. 1 is that one must generate two tightly controlled clocks off-chip, with the resolution limited by the jitter with which these clocks can be generated. Instead, we wish to generate the trigger and sample edges on chip, derived from the same clock reference, with the interval between them determined by digital control, a digital-to-time converter. The circuits required are similar to those employed in time-to-digital converters [14]. The simplest way to build a digital-to-time converter is with a delay-locked loop (DLL) as shown in Fig. 6. In this case, the N -stage voltage-controlled delay line (VCDL) is locked to one period (T_{clk}) of the reference clock. This gives each buffer stage

⁵The clock fall time is approximately 100 ps. The clock delays to the nFET and pFET of the complementary switch are carefully matched.

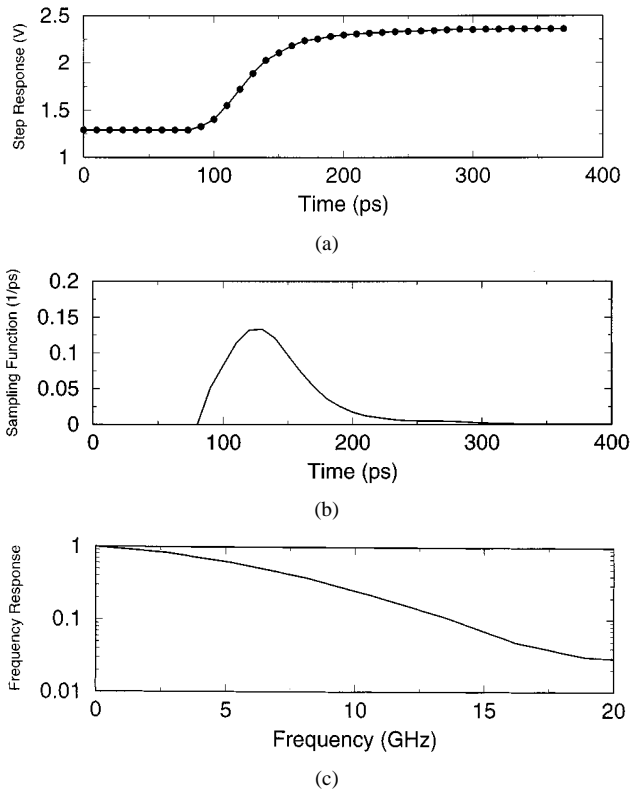


Fig. 5. (a) Step response. (b) Impulse response of the sampler. (c) Frequency response of the switch, determined from the Fourier transform of the impulse response.

of the VCDL a precise delay of T_{clk}/N .⁶ By multiplexing out the outputs of the buffer stages, one could create sample and trigger edges separated by multiples of T_{clk}/N .

The use of a single DLL, however, limits the time resolution to a gate delay. One technique for overcoming this would be to introduce a circuit to interpolate between the delay stages [15]. Instead, we decided on a Vernier approach using two DLLs as shown in Fig. 7. In this case, one DLL has a VCDL with N stages and the other has a VCDL with M stages, both locked to the same reference clock. In this case, the delay of each buffer in the first VCDL is locked to a delay of T_{clk}/N and the delay of each buffer in the second VCDL is locked to a delay of T_{clk}/M . By choosing the sample clock from one DLL and the trigger clock from the other, one can achieve multiples of a timing resolution of $T_{\text{clk}}/N - T_{\text{clk}}/M$, which can be a fraction of a gate delay.

The DLLs used in this design will be embedded in a hostile digital environment. As such, they must be as immune as possible to jitter caused by substrate and power-supply noise. To accomplish this, the VCDL is constructed with differential buffers as shown in Fig. 8 with “symmetric” loads defined by a diode-connected pFET (with a diode-like characteristic) in parallel with a biased pFET (with a triode-like characteristic) [16]. The opposite curvatures of the two characteristics combine to produce a nearly linear load, limiting the conversion of common-mode supply noise into differential jitter. In addition,

⁶This assumes that the buffer stages are perfectly matched. Any static phase offsets in the phase detector and charge pump will also be spread across the buffer stages.

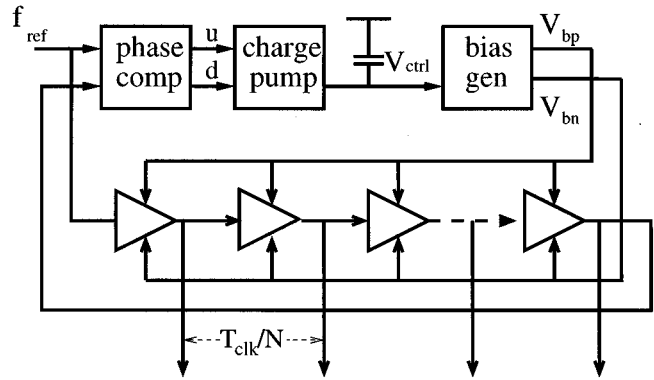


Fig. 6. A delay-locked loop can be used to generate well-controlled delay intervals based on a reference clock. The digital-to-time converters use DLLs based on the design of [17].

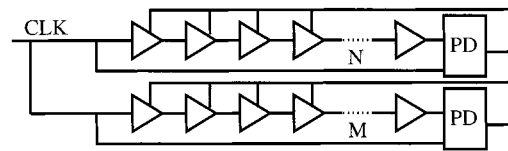


Fig. 7. Two delay-locked loops can be used to achieve a timing resolution of $T_{\text{clk}}/N - T_{\text{clk}}/M$ by choosing the sample clock from one DLL and the trigger clock from the other.

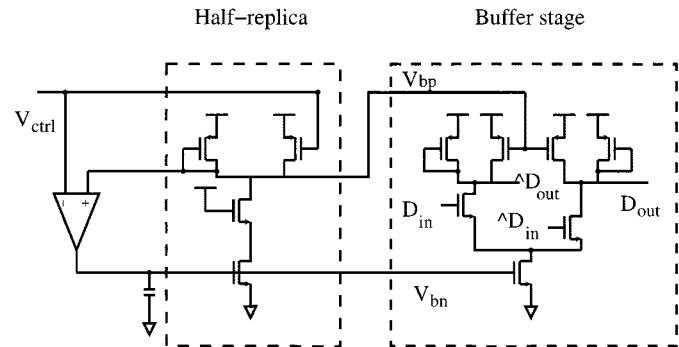


Fig. 8. The buffers of the DLL are self-biased by a half-replica of the differential-pair buffer, locking the lower limit of the output swing to the control voltage.

the buffers are self-biased by a half-replica of the differential pair, locking the lower limit of the output swing to the control voltage V_{ctrl} [17]. There are stability issues associated with this control loop. The loading at the output of the differential amplifier must be sufficient to produce dominant-pole compensation and an overall phase margin of at least 35° , but the loading cannot be so large as to reduce the closed-loop bandwidth excessively and limit dynamic power-supply noise rejection [18]. A loading of about ten buffer stages per bias generator is an appropriate compromise in our case.

The digital-to-time converter on our test chip combines two DLLs, one with 30 buffers and the other with 32 buffers with a 200-MHz reference clock ($N = 30$, $M = 32$). The buffer stages are carefully matched and the outputs of the buffers are multiplexed to produce time separations between the trigger and sample clock in steps of 10.4 ps ($T_{\text{clk}}(1/N - 1/M)$) up to 256 steps or 2.5 ns ($T_{\text{clk}}/2$) as shown in Fig. 9.

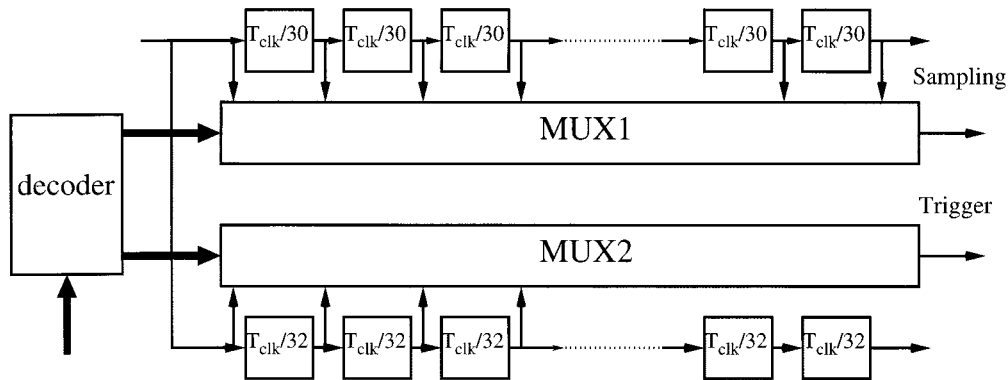


Fig. 9. Digital-to-time converter combining two DLLs locked to the same reference clock.

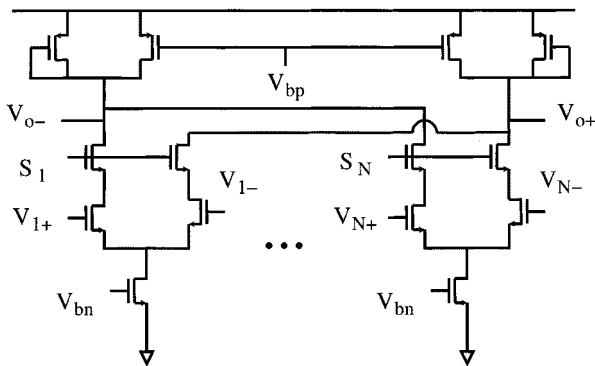


Fig. 10. Schematic of the multiplexers.

The decode logic of Fig. 9 converts an 8-bit input address into 256 different delays between the sample clock and trigger clock. Let $D_1 = T_{\text{clk}}/30$ be the stage delay for the 30-stage DLL. Let $D_2 = T_{\text{clk}}/32$ be the stage delay for the 32-stage DLL. The resolution is then $d = D_1 - D_2$ with $D_1 = 16d$ and $D_2 = 15d$. Let a_0 be the four least significant bits of the address and a_1 be the four most significant bits of the address. Then the sample clock is chosen by decoding and multiplexing (through MUX1) $a_0 + a_1$ and the trigger clock is chosen by decoding and multiplexing (through MUX2) a_0 . MUX1 is then choosing one of the 31 outputs of the VCDL as shown in Fig. 9. MUX2 is choosing one of the outputs of the first 16 delay stages of the 32-stage DLL. The sample clock delay in this case is given by $(a_0 + a_1)D_1$ and the trigger clock delay is given by a_0D_2 . The difference is $a_0d + 16a_1d$, which is a direct “decode” of the digital word into multiples of d .

The schematic of a multiplexer is shown in Fig. 10 [16], where S_1 to S_N are the select signals for different stages; (V_{1+}, V_{1-}) to (V_{N+}, V_{N-}) are the multiplexed outputs of those buffer stages. The differential outputs of the multiplexer (V_{o+}, V_{o-}) are converted to full-rail by a differential-to-single-ended converter. Based on the buffer stage in Fig. 8, both the multiplexers and the differential-to-single-ended converters are also controlled by V_{bn} and V_{bp} generated by bias generator to minimize power-supply noise sensitivity. The interconnects for the stages are carefully matched in the layout.

We note that the DLLs used in our digital-to-time converter are larger (the area consumed by the two DLLs exceeds 0.05 mm^2) than necessary. With the availability of a PLL-gen-

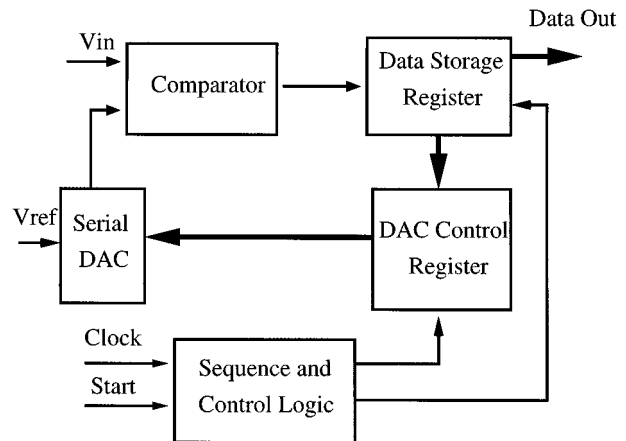


Fig. 11. A simple successive approximation ADC design is used in the scope.

erated high-frequency on-chip clock, fewer stages could be used in the delay lines. Additionally, we have sized the differential delay buffers in the VCDL conservatively to improve matching and reduce “side-branch” loading mismatch from the multiplexers.

V. ANALOG-TO-DIGITAL CONVERTER

The 8-bit ADC uses a successive-approximation (SA) algorithm and a two-capacitor serial DAC [19] as shown in Fig. 11. The capacitors in the serial DAC are implemented using metal-insulator-metal (MIM) capacitors between *metal4* and a special metal layer, giving a capacitance of $1 \text{ fF}/\mu\text{m}^2$. The comparator design is shown in Fig. 12 [20]. In track mode, the comparator has a gain of approximately 21.9 dB with the gain around the positive feedback loop shunted to be less than one (ensuring stability). In latch mode, the regenerative action is enabled, producing nearly full-rail output. This track-and-latch architecture gives good comparator resolution without the need for a multistage amplifier. The overall SA ADC design, though slow, is fairly area-efficient, consuming less than 0.01 mm^2 .

VI. TEST CHIP DESIGN

The overall design of the test chip is shown in Fig. 13. The control logic steps the digital-to-time converter through increments of 10.4 ps, from a user-specified start time to a user-speci-

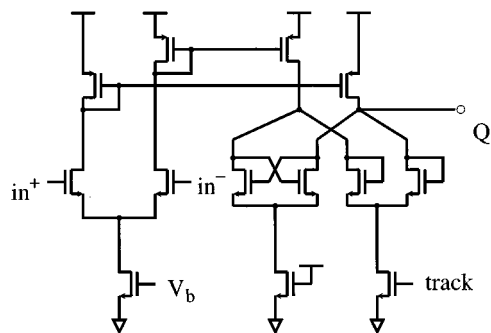


Fig. 12. Track-and-latch comparator used in the ADC.

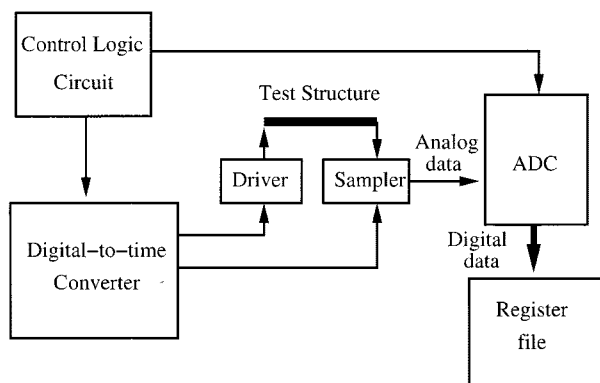


Fig. 13. Overall test chip design, incorporating the sample, digital-to-time converter, and ADC.

fied end time. All of the samples are stored in a 2048-bit register file, which can be scanned out after measurement completion.

In general, there can be multiple samplers that can share all the rest of the oscilloscope circuitry, keeping the overhead low for a large digital chip. In our design, seven “sampling heads” are multiplexed to the scope. Large sampler fan-in can be easily accommodated because of the “slow” time scales of the data conversion of the sampled voltage.

The test chip was designed in the TSMC 0.25- μm 5M1P process. This is a 2.5-V process with transistor saturation currents at maximum overdrive of about $600 \mu\text{A}/\mu\text{m}$ for the nFET and $300 \mu\text{A}/\mu\text{m}$ for the pFET. There are five levels of AlCu interconnect. The first four levels have sheet resistivities of $0.076 \Omega/\square$. *metal5* has a sheet resistivity of $0.044 \Omega/\square$. A die photo of the fabricated test chip is shown in Fig. 15. Seven samplers, with the circuit schematic of Fig. 4, are positioned to measure various waveforms within a snaking 4-mm-long 16-bit bus structure. A more detailed picture of the interconnect structure is shown in Fig. 14. This structure is evident in the top, left corner of the chip (see Fig. 15). The spacing of the power-ground grid is $100 \mu\text{m}$ and this is routed horizontally in *metal2* and *metal4* and vertically in *metal3* and *metal5*. The bus is routed within this grid on *metal5* vertically and *metal4* horizontally. Samplers are placed on the far end and near end of bits 0, 3, and 7 of the bus and on the trigger signal. The drivers of the bus are designed to switch with one of three strengths or hold the net high or low. 4 pF of thin-oxide on-chip decoupling capacitance is used to minimize power-supply noise created by the switching buffers. The receiver loads are also variable

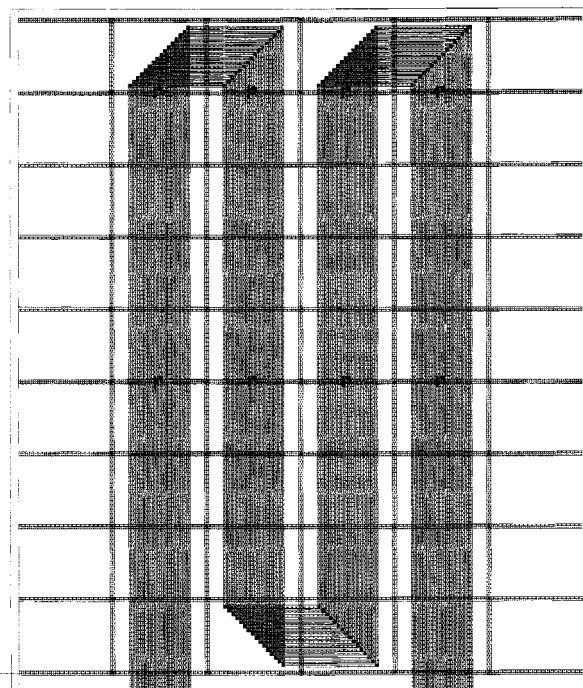


Fig. 14. Test structure of test chip: a 16-bit bus snakes through the power grid.

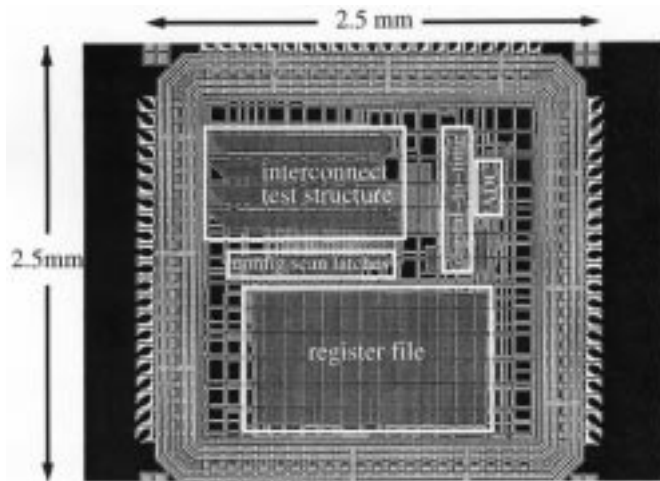


Fig. 15. Die photo of test chip.

with MOS switches determining variable amounts of MOS capacitance that can be added to the far-end. The configuration of the test site is determined by a set of scan-only flip-flop which set the driver and receiver configurations and enable one of the samplers.

VII. RESULTS

The test chip layout was extracted (resistances, capacitors and inductors) using Cadence’s Assura RCX-PL extraction engine [2] and simulated with HSPICE. These simulation results can be compared with the time-domain measurement results on an absolute time scale because we are also sampling the trigger clock (and, therefore, know the measured time points relative to the trigger clock).

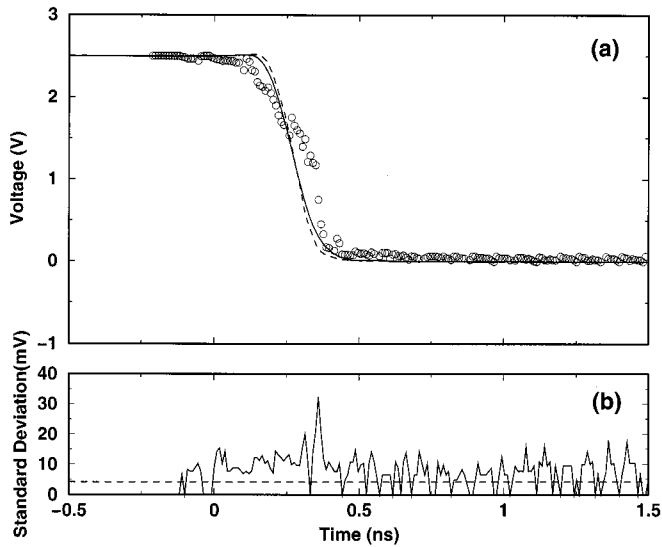


Fig. 16. Measured near-end data on bit 7 with all the bits of the bus switching simultaneously: (a) Circles represent the average of 20 identical measurements. The solid curve is the simulation result with the *RLCK* netlist extracted by Assura RCLX, convolved with the sampling function. The dashed curve is the *RC*-only simulation result. (b) Standard deviation in the measured data.

A few comments on the parasitic extraction used in our HSPICE simulations are in order. The extraction approach used is that of return-limited inductance extraction [2], [21], which produces a coupled *RLCK*⁷ netlist for the signal lines, assuming uniform current density across the wire cross section (no skin effect). This extraction approach treats the substrate and the power-ground grids as ideal equipotentials. The substrate is, furthermore, assumed to be too resistive for eddy currents to be induced (i.e., it is ignored in the magnetostatic inductance calculation). Return-limited inductance extraction further limits inductive coupling to an interaction region around the wires defined by the nearest power-ground lines.

The circles in Figs. 16(a) and 17(a) are the actual measured data (near-end and far-end, respectively) on a switching bit 7 (in the middle of the bus) in the presence of simultaneous switching on the other 15 bits of the bus. The measured results represent the average of 20 measurements done under identical conditions. The strongest drive strength and minimum receiver load capacitance are configured for the measurement. Simultaneous switching is of interest because it “boosts” the effective inductance of bit 7 by the mutual inductances to the other 15 bits of the bus. The simulation results are presented on the same graphs for comparison. The solid curve is the HSPICE result from *RLCK* extracted data convolved with the sampling function of Fig. 5(b) to consider the effects of finite sample bandwidth. The finite bandwidth of the sampler has only a modest effect on this waveform. The dashed curves are the result of HSPICE simulation in the absence of inductance (i.e., an *RC*-only interconnect model). Clear ringing is observed in the far-end waveform (Fig. 17) in both measurement and simulation. The measured data show a clear voltage drop before the actual switching transition, which we attribute to power-supply noise due to the action of the predrivers preceding the large buffers used to drive the

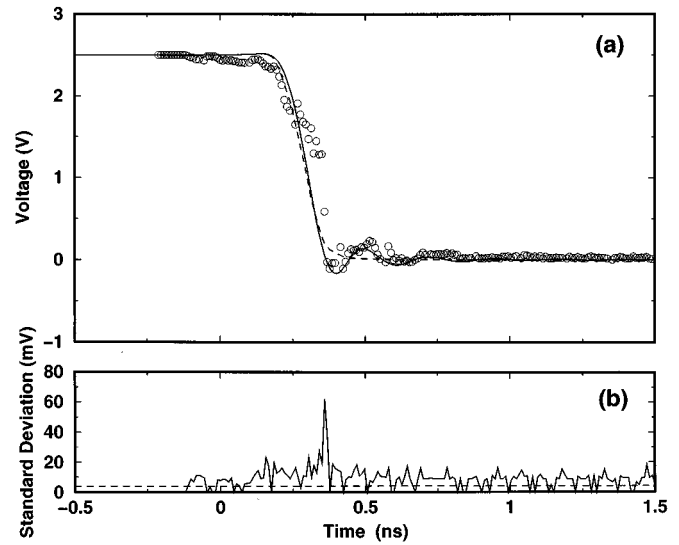


Fig. 17. Measured far-end data on bit 7 with all the bits of the bus switching simultaneously: (a) Circles represent the average of 20 identical measurements. The solid curve is the simulation result with the *RLCK* netlist extracted by Assura RCLX, convolved with the sampling function. The dashed curve is the *RC*-only simulation result. (b) Standard deviation in the measured data.

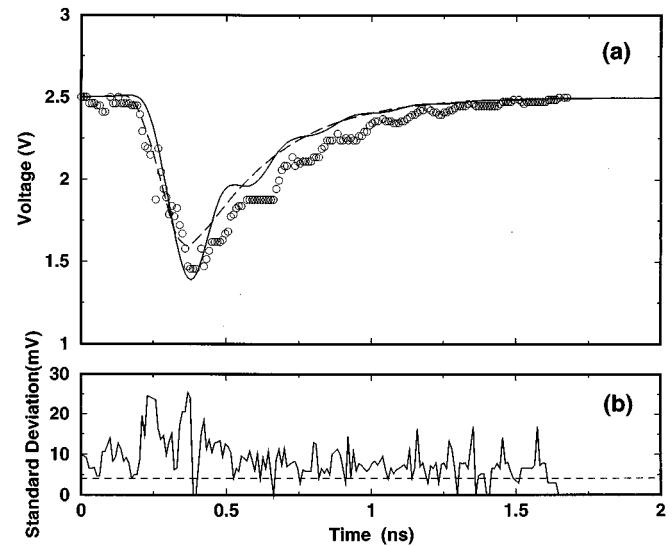


Fig. 18. Measured crosstalk on bit 7 (far-end) due to the simultaneous switching of all of the other bits of the bus: (a) Circles represent the average of 20 identical measurements. The solid curve is the simulation result with the *RLCK* netlist extracted by Assura RCLX, convolved with the sampling function. The dashed curve is the *RC*-only simulation result. (b) Standard deviation in the measured data.

bus. These buffers switch and introduce “droop” in the power supply slightly before the transition of the bus structure. (Power-supply noise is not modeled in our analysis, since the power supply is assumed to be a rigid equipotential in our extraction.)

Crosstalk noise due to the switching of all the other bits while bit 7 is quiet is shown in Fig. 18. Simulation results from *RLCK* and *RC* extract are once again presented as the solid and dashed curves, respectively. There is clear ringing evident in both the measured and simulated result on the “trailing” edge of the crosstalk noise waveform. Ringing seems slightly more pronounced in the measured results than in simulation.

⁷By “*K*”, we mean the normalized mutual inductance $K = M/\sqrt{L_1 L_2}$.

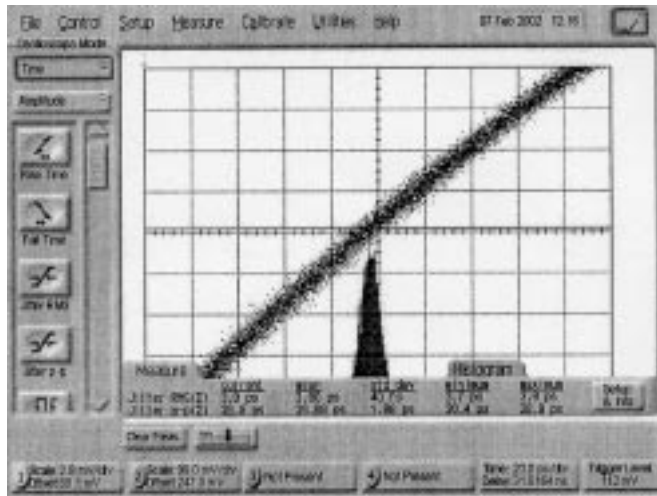


Fig. 19. Jitter histogram of the “trigger” DLL output in the absence of switching activity on the 4-mm bus.

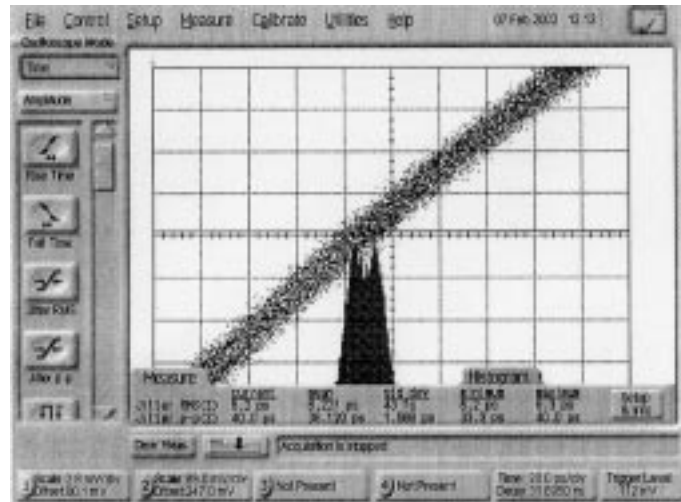


Fig. 20. Jitter histogram of the “trigger” DLL output in the presence of switching activity on the 4-mm bus.

We attribute much of the “noise” observed in the measured results to errors in the generation of the sample and trigger edges since this noise seems most pronounced when the sample voltage is rapidly changing. The timing resolution of the digital-to-time converter is limited by the jitter of the DLLs as well as by error in the (static and dynamic) matching of the buffer stages of the VCDL.

To understand these noise issues better, particularly in the presence of the power-supply noise introduced by the switching drivers, we performed two experiments—we externally measured the jitter of the DLLs and through the multiple measurements, we determined the variance in the measured waveform data. Fig. 19 shows the jitter histogram of the DLL output buffered off-chip in the absence of switching activity in the test site (i.e., switching of the large drivers on the 4-mm bus) as measured by an Agilent 86100A wide-bandwidth sampling oscilloscope. The 200-MHz external reference clock is generated by an Agilent 81130A pulse/data generator. This clock has a cycle-to-cycle jitter (i.e., the jitter in the period of the output) of approximately 3.5 ps rms (22.7 ps peak-to-peak). The cycle-to-cycle jitter of the buffered DLL output is 3.9 ps rms (28.9 ps peak-to-peak). By contrast, Fig. 20 shows the jitter histogram in the presence of switching activity on the 4-mm bus, as will be the case when the oscilloscope is operating. In this case, the rms jitter has increased to 6.2 ps (40.0 ps peak-to-peak). Subtracting (in an rms way) the measured jitter of the reference clock from the DLL output jitter with bus switching activity yields an rms jitter introduced by each DLL of approximately $t_{\text{jitter}}^{\text{rms}} = 4.8$ ps.⁸ Assuming that this jitter is introduced primarily by the VCDL, each stage contributes $1/N$ to the total rms jitter for an N -stage DLL. Assuming that the jitter of the sample and trigger DLLs is uncorrelated and using the definitions of a_0 , a_1 , N , and M introduced in Section IV,

the relative jitter between the trigger and sample clocks is dependent on the chosen interval and given by

$$\left(\sqrt{\frac{(a_0 + a_1)^2}{N^2} + \frac{a_0^2}{M^2}} \right) t_{\text{jitter}}^{\text{rms}}.$$

This is maximum ($\cong 1.1t_{\text{jitter}}^{\text{rms}}$) for the $T_{\text{clk}}/2$ interval. Additional jitter is likely introduced by the multiplexers and clock buffering that is not included in this analysis. Since the waveforms were repeated 20 times, the associated standard deviations can be calculated and are shown in Figs. 16(b), 17(b), and 18(b). The quantization noise floor of the 8-b DAC (calculated as $V_{\text{lsb}}/\sqrt{12}$) is noted for reference. The standard deviations are clearly peaked in the regions of rapid voltage change.

Artifacts clearly remain in the average waveforms indicating the presence of correlated power-supply noise and offset errors. Some of the jitter measured in the DLL may actually be deterministic in relation to the measured waveforms (and will, therefore, appear as offset), since the power-supply noise from the switching bus structure has a correlated relationship to the sampling time.

VIII. CONCLUSION AND APPLICATIONS

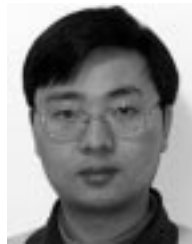
In this paper, we have described the first self-contained, on-chip sampling oscilloscopes for the measurement of high-speed analog waveforms in digital and mixed-signal integrated circuits. The chip employs subsampling techniques enabled by an on-chip digital-to-time converter with (ideally) 10-ps resolution. Eight-bit digital data from an area-efficient successive-approximation ADC is stored in a scannable register file.

To employ this technique within the design-for-testability (DFT) methodology of a digital integrated circuit, samplers would have to be positioned near each critical net “tap” point. The digital-to-time converter and the ADC can be shared across all of the samplers and can be positioned anywhere on the chip.

⁸Some of the measured DLL jitter could have also been introduced by the drive path to get the clock off-chip. We are conservatively assigning it all to the DLL.

REFERENCES

- [1] G. W. Roberts, "Improving the testability of mixed-signal integrated circuits," in *Proc. Int. Custom Integrated Circuits Conf.*, 1997, pp. 214–221.
- [2] K. L. Shepard, D. Sitaram, and Y. Zheng, "Full-chip, three-dimensional, shapes-based RLC extraction," in *Proc. Int. Conf. Computer-Aided Design*, 2000, pp. 142–149.
- [3] W. R. Eisenstadt and Y. Eo, "S-parameter based IC interconnect transmission line characterization," *IEEE Trans. Comp., Hybrids, and Manufact. Technol.*, vol. 15, no. 4, pp. 483–490, 1992.
- [4] B. Kleveland, C. H. Diaz, D. Vook, L. Madden, T. H. Lee, and S. S. Wong, "Exploiting CMOS reverse interconnect scaling in multigigahertz amplifier and oscillator design," *IEEE J. Solid-State Circuits*, vol. 36, pp. 1480–1488, Oct. 2001.
- [5] A. Deutsch, G. V. Kopsay, C. W. Surovic, B. J. Rubin, L. M. Terman Jr., R. P. Dunne, T. A. Gallo, and R. H. Dennard, "Modeling and characterization of long on-chip interconnections for high-performance microprocessors," *IBM J. Res. Develop.*, vol. 39, no. 5, pp. 547–567, 1995.
- [6] K. Soumyanath, S. Borkar, C. Zhou, and B. A. Bloechel, "Accurate on-chip interconnect evaluation: a time-domain technique," *IEEE J. Solid-State Circuits*, vol. 34, pp. 623–631, May 1999.
- [7] P. Larsson and C. Svensson, "Measuring high-bandwidth signals in CMOS circuits," *Electron. Lett.*, vol. 29, no. 20, pp. 1761–1762, Sept. 1993.
- [8] K. Lofstrom, "Early capture for boundary scan timing measurement," in *Proc. IEEE Int. Test Conf.*, Oct. 1996, pp. 417–422.
- [9] M. A. Burns, "Understampling Digitizer With a Sampling Circuit Positioned on an Integrated Circuits," U. S. Patent 5 578 935, 1996.
- [10] A. Hajjar and G. W. Roberts, "A high speed and area efficient-on-chip analog waveform extractor," in *Proc. IEEE Int. Test Conf.*, Oct. 1998, pp. 688–697.
- [11] R. Ho, B. Amrutur, K. Mai, B. Wilburn, T. Mori, and M. Horowitz, "Applications of on-chip samplers for test and measurement of integrated circuits," in *Proc. Symp. VLSI Circuits Dig. Tech. Papers*, 1998, pp. 138–139.
- [12] K. L. Shepard and Y. Zheng, "On-chip oscilloscopes for noninvasive time-domain measurement of waveforms," in *Proc. Int. Conf. Comput. Design*, 2001, pp. 221–226.
- [13] H. O. Johansson and C. Svensson, "Time resolution of NMOS sampling switches used on low-swing signals," *IEEE J. Solid-State Circuits*, vol. 33, pp. 237–245, Feb. 1998.
- [14] J. Christiansen, "An integrated high resolution CMOS timing generator based on an array of delay locked loops," *IEEE J. Solid-State Circuits*, vol. 31, pp. 952–957, July 1996.
- [15] C.-K. K. Yang and M. A. Horowitz, "A 0.8 μm CMOS 2.5 Gb/s oversampling receiver and transmitter for serial links," *IEEE J. Solid-State Circuits*, vol. 31, pp. 2015–2023, Dec. 1996.
- [16] J. G. Maneatis and M. A. Horowitz, "Precise delay generation using coupled oscillators," *IEEE J. Solid-State Circuits*, vol. 28, pp. 1273–1282, Dec. 1993.
- [17] J. G. Maneatis, "Low-jitter process-independent DLL and PLL based on self-biased techniques," *IEEE J. Solid-State Circuits*, vol. 31, pp. 1723–1732, Nov. 1996.
- [18] W. J. Dally and J. W. Poulton, *Digital Systems Engineering*. Cambridge, MA: Cambridge Univ. Press, 1998.
- [19] R. E. Suarez, P. R. Gray, and D. A. Hodges, "All-MOS charge redistribution analog-to-digital conversion techniques—part II," *IEEE J. Solid-State Circuits*, vol. SC-10, pp. 379–385, Dec. 1975.
- [20] B. Song, H. S. Lee, and M. Tompsett, "A 10-b 15-MHz CMOS recycling two-step A/D converter," *IEEE J. Solid-State Circuits*, vol. 25, pp. 1328–1338, Dec. 1990.
- [21] K. L. Shepard and Z. Tian, "Return-limited inductances: a practical approach to on-chip inductance extraction," *IEEE Trans. Computer-Aided Design*, vol. 19, pp. 425–436, Apr. 2000.



Yu Zheng (S'01) received the B.S. and M.S. degrees from the Department of Computer Science and Technology at Peking University, Beijing, China, in microelectronics, in 1996 and 1999, respectively, the Ph.D. degree from Columbia University, New York in 2003.

He is currently with Cadence Design Systems, San Jose, CA.



Kenneth L. Shepard (S'85–M'92) received the B.S.E. degree from Princeton University, Princeton, NJ, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Palo Alto, CA, in 1987, 1988, and 1992, respectively.

From 1992 to 1997, he was a Research Staff Member and Manager in the VLSI Design Department, IBM T. J. Watson Research Center, Yorktown Heights, NY, where he was responsible for the design methodology for IBM's G4 S/390 microprocessors. Since 1997, he has been at Columbia University,

New York, NY, where he is currently an Associate Professor. He also served as Chief Technology Officer of CadMOS Design Technology, San Jose, CA, until its acquisition by Cadence Design Systems in 2001. His current research interests include design tools for advanced CMOS technology, on-chip test and measurement circuitry, low-power design techniques for digital signal processing, low-power intrachip communications, and CMOS imaging applied to biological applications.

Dr. Shepard received the Fannie and John Hertz Foundation Doctoral Thesis Prize in 1992. At IBM, he received Research Division Awards in 1995 and 1997. He was also the recipient of an National Science Foundation CAREER Award in 1998 and the IBM University Partnership Awards in 1998, 1999, 2000, and 2001. He was also awarded the 1999 Distinguished Faculty Teaching Award from the Columbia Engineering School Alumni Association. He has been an Associate Editor of IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION SYSTEMS and was the Technical Program Chair for the 2002 International Conference on Computer Design. He has served on the program committees for DAC, ICCAD, ISCAS, ISQED, GLS-VLSI, TAU, and ICCD.