An Alternate Design Paradigm for Low-Power, Low-Cost, Testable Hybrid Systems using Scaled LTPS TFTs

JING LI
Purdue University
ADITYA BANSAL
IBM T.J. Watson Research
SWAROP GHOSH
Purdue University
and
KAUSHIK ROY
Purdue University

13

This article presents a holistic hybrid design methodology for low-power, low-cost, testable digital designs using low-temperature polycrystalline-silicon thin-film transistors (LTPS TFTs). An alternate scaling rule under low thermal budget (due to flexible substrate) is developed to improve the performance of TFTs in the presence of process variation. We demonstrate that LTPS TFTs can be further optimized for ultralow-power subthreshold operation with performances comparable to contemporary single-crystal silicon-on-insulator (c-Si SOI) devices after process optimization. The optimized LTPS TFTs with high current drivability and less variability can comprise a promising low-cost option to augment Si CMOS technology, opening up a plethora of new hybrid 3D applications. We illustrate one such application: IC testing. Testing of complex VLSI systems is a prime concern due to design cost of DFT circuits, area/delay overheads, and poor test confidence. To harness the benefits of TFT technology, a novel low-power, process-tolerant, generic, and reconfigurable test structure designed using LTPS TFTs is proposed to reduce the test cost, as well as to improve diagnosability and verifiability, of complex VLSI systems. Due to proper optimization of TFT devices, the proposed test structure consumes low power but operates with reasonable performance. Furthermore, the test circuits do not consume any silicon area because they can be integrated on-chip using 3D technology. Since the test architecture is reconfigurable, this

This research was supported by DARPA.
Authors' addresses: J. Li, Department of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907-1285; email: jingli@purdue.edu; A. Bansal, IBM T.J. Watson Research Center, 1101 Kitchawan Rd., Yorktown Heights, NY 10598; email: bansal@us.ibm.com; S. Ghosh, K. Roy, Department of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907-1285; email: {ghosh3, kaushik}@purdue.edu.

eliminates the need to redesign built-in-self-test (BIST) components that may vary from one processor generation to another. We have developed test structures using 200nm TFT devices and evaluated them on designs implemented in 130nm bulk CMOS. For circuit simulations, we have developed a SPICE-compatible model for TFT devices. The BIST components designed using the test structures operate at 0.8–4.3 GHz (compared to 8.2 GHz in bulk CMOS) with low power consumption. The enhanced scan cells partially implemented in TFT (3D hybrid design) consume ∼24% less power and ∼15–20% less area of Si die compared to conventional bulk-Si design (2D planar design), with minimal delay overhead.

Categories and Subject Descriptors: B.7.3 [**Integrated Circuits**]: Reliability and Testing; D.2.2 [**Software Engineering**]: Design Tools and Techniques

General Terms: Design, Performance, Reliability

Additional Key Words and Phrases: Low-temperature polycrystalline silicon (LTPS), thin-film transistor (TFT), grain boundary (GB), hybrid system, 3D integration, inherent variation, generic, reconfigurable, DFT, BIST

ACM Reference Format:

Li, J., Bansal, A., Ghosh, S., and Roy, K. 2008. An alternate design paradigm for low-power, low-cost, testable hybrid systems using scaled LTPS TFTs. ACM J. Emerg. Technol. Comput. Syst. 4, 3, Article 13 (August 2008), 19 pages. DOI = 10.1145/1389089.1389093 http://doi.acm.org/10.1145/1389089.1389093

1. INTRODUCTION

Continuous scaling of transistor dimensions due to advances in IC technology (e.g., lithography and etching) allows us to put more functionality on a chip. However, aggressive scaling in CMOS-feature size also creates several challenging problems, such as interconnect delay, power consumption, die overheating, and functional/delay failures. With increasing integration density, test/verification of such complex systems also becomes a challenging task. Hence, it is extremely difficult to carry ICs further along the path of Moore's law beyond 32nm technologies. Recently, researchers have shown that scaling in the vertical dimension (i.e., 3D integration) can be a promising solution to mitigate interconnect delays and to put more functionality on the chip [Strickland et al. 1998; Patti 2006]. In contrast to conventional 2D planar chips, this technique offers significant improvement in performance and reduction in power dissipation [Patti 2006]. However, the associated fabrication (i.e., SOI technology) and integration challenges (i.e., optical alignment) may substantially increase the total cost [Patti 2006].

To ensure that 3D ICs are not only technologically but also economically viable, it is imperative to explore less expensive, alternate technologies that may open up different application possibilities while making systems more reliable and testable. One such emerging technology is that using low-temperature polycrystalline-silicon thin-film transistors (LTPS TFTs). In contrast to bulk-CMOS technology, LTPS TFT has much lower fabrication cost due to: (a) low thermal budget, (b) low masking cost, and (c) the flexibility of being fabricated on a wide range of substrates such as glass, plastic, and so on. Furthermore, fabrication on transparent substrates such as glass leads to a much simpler alignment process in 3D integration as compared to contemporary CMOS-CMOS integration. In spite of the aforementioned advantages, state-of-the-art LTPS

TFT technology has limited applications; it is used mainly for liquid crystal displays (LCDs). This is primarily due to its lower current drivability. Poor device characteristics require high supply voltages (10–20V) to meet the performance target [Tanaka et al. 1993], resulting in high power consumption. It has been experimentally demonstrated and analytically verified that the limiting factor in LTPS TFT performance originates from the channel material [Walker et al. 2004]. In contrast to bulk MOSFETs, the polycrystalline-silicon channel material in LTPS TFTs consists of a number of single-crystal grains with highly defective regions in between, called grain boundaries (GBs) [Walker et al. 2004]. Such GBs are randomly oriented and distributed in the channel region, inherently restricting the performance of the device. For conventional LCD applications, the transistor-channel length is typically large ($>10\mu$m) and therefore a large number of GBs can be present in the channel. As a consequence, the on-current (I_{on}) of the device is severely degraded, making long-length transistor-channel devices unsuitable for high-performance applications. In Karaki et al. [2005], the authors proposed an asynchronous circuit design technique to mitigate the process variation in LTPS TFTs. However, asynchronous design requires fundamental modification of the well-proven synchronous design flow, resulting in high design overhead. To successfully implement TFTs in synchronous systems will require further exploration into device/circuit innovations for low-power and variation resilience.

To harness the true benefits of TFTs, it is essential to overcome their drawbacks, namely poor on-current and grain boundary (GB)-induced variation. In this article, we achieve this target by means of an alternate design paradigm: device/circuit/architecture codesign. We optimize LTPS TFTs and make them compatible with CMOS and suitable for 3D hybrid systems. From a device perspective, a novel scaling rule under low thermal budget (due to flexible substrate) is proposed to keep the number of GBs under control, thus improving the performance. Device optimization is performed both at the design phase and the manufacturing phase. We also demonstrate that such an approach can be promising for ultralow-power subthreshold operation, with performance comparable to contemporary c-Si SOI devices [Li et al. 2007a].

However, for TFTs, statistical variations induced by GBs are a major concern [Li et al. 2007b, 2007c] and are more significant than other parametric variations (i.e., channel length and width, and oxide thickness). To ensure robust and stable functionality of TFT circuits, we developed an efficient variability compensation technique, namely multifinger structure, to aggressively reduce the variation. Such LTPS TFTs with high current drivability and less variability can comprise a promising candidate for 3D hybrid systems [Li et al. 2007d]. To demonstrate the feasibility of such hybrid technologies, we integrate test structures designed using TFTs with underlying CMOS circuits, to create highly reliable systems. Note that our test approach is very different from the CMOS-CMOS 3D stacking technique [Mysore et al. 2006] that has been used for chip analysis (by monitoring the hardware performance of the microprocessor). Although Mysore's technique improves verification and diagnosis capability of the system, its application is limited to online monitoring and neither generic nor reconfigurable. Furthermore, it does not address one of the major issues

associated with offline test, namely design-for-test overhead, that is the focus of this work. This fact will be elaborated in Section 3.

In particular, we have made following contributions in this article.

1. We have developed a device optimization methodology for LTPS TFTs operating at CMOS-compatible supply voltages, with low power consumption and high performance.
2. We have developed a statistical simulation methodology based on device material properties to illustrate the inherent variations in scaled LTPS TFT technology. An efficient variability compensation technique, namely multi-finger structure, is proposed to reduce the effects of variation.
3. We have proposed a hybrid system using low-cost LTPS TFTs for offline as well as online test of underlying complex VLSI systems. The test architecture is generic and fully reconfigurable, and it eliminates the need to redesign BIST components that may vary from one processor generation to another. The test structures designed using LTPS TFTs significantly reduce the cost as compared to conventional Si CMOS 3D technology.

The rest of the work is organized as follows. In Section 2, we explain the basic device physics of LTPS TFTs and the optimization methodology for low power and high performance in the presence of process variation. The hybrid system for testability using the proposed LTPS TFT technology is introduced in Section 3. The architecture of TFT test structure and the overall test approach are also described in that section. The practical challenges and issues are addressed in Section 4, followed by conclusions in Section 5.

2. DEVICE/CIRCUIT OPTIMIZATION STRATEGY

For creating a hybrid system, the operating voltages of the TFT and the bulk CMOS should be compatible in order to avoid level converters. Standard applications of TFTs, such as LCDs, usually operate at 10–20V. Since TFTs are low-performance devices, high supply voltages are typically required to increase the drive current. High operating voltages inhibit the scaling of gate oxide thickness. In this section, we present a device optimization methodology to achieve sufficient current drivability under CMOS-compatible power supply voltages. We also provide practical insights into circuit design with TFTs and discuss the associated challenges.

2.1 Preliminaries: Grain Boundaries (GBs) in LTPS TFTs

LTPS (or simply poly-Si) TFTs have similar device structure as the single-crystal silicon-on-insulator (c-Si SOI) devices, except that the channel material is polycrystalline silicon, not single-crystalline silicon. Poly-Si material consists of a number of single-crystal grains with highly defective regions in between, called grain boundaries (GBs). A GB can trap free-charge carriers, making them immobile. As a consequence, the effective number of carriers available for electrical conduction is reduced and the GB region gets negatively/positively charged (depending on trapping electrons or holes), resulting in reduced drive current. To keep the charge neutrality, positively/negatively charged depletion

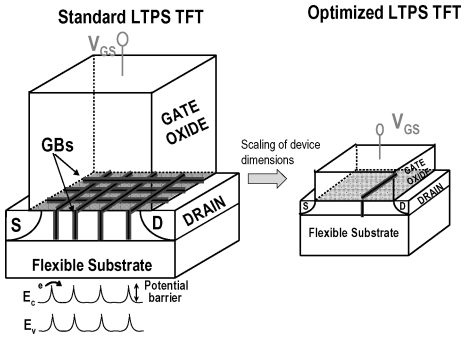

Fig. 1. Schematic showing standard LTPS TFT and optimized LTPS TFT. The number of GBs is reduced significantly.

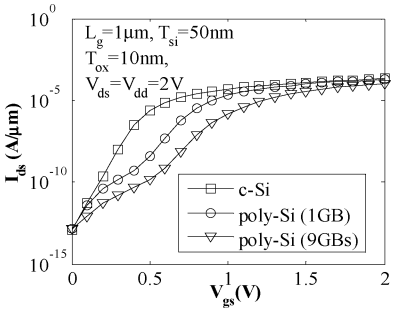

Fig. 2. Ids-Vgs curves of long-channel TFT with different numbers of GBs. A reduced number of GBs improves the IV characteristics.

regions are formed at both sides of the GB. The potential barrier due to the charge-trapping mechanism impedes the motion of carriers from one grain to another, which further reduces the conductivity.

In macroelectronic applications (such as LCDs), the transistor-gate length is typically large ($\gg 10\mu$m) and therefore a large number of GBs can be present in the channel (Figure 1). As a consequence, the on-current (I_{on}) of the device degrades considerably. To achieve sufficient current drivability, supply voltage has to be increased significantly (10–20V). Under such great dynamic stress, the gate dielectric thickness (T_{ox}) should also be large enough to achieve high reliability. Hence, such transistors are not suitable for high-performance and/or low-power circuits.

2.2 Design Perspective: Technology Scaling under Low Thermal Budget

As discussed in the previous section, GBs are the limiting factor for device performance (Figure 2). Hence, to improve the performance, there are two possible methods: (a) increase the grain size, thereby reducing the number of GBs; or (b) scale down the device size so that the number of GBs in the channel is reduced and I_{on} is improved. Enlarging the grain size (option (a)) can be accomplished with complex processes such as metal-induced lateral

Table I.

Device Characteristics	c-Si SOI	poly-Si TFT	Small grain poly-Si TFT
Printed Gate Length, L (nm)	200	200	200
Average Grain Size, L_{gr} (nm)	–	200	50
Oxide Thickness, T_{ox} (nm)	3.2	10	10
Si Film Thickness, T_{si} (nm)	35	6-20	6-20

crystallization (MILC) [Wang et al. 2001], solid-phase crystallization (SPC) [Subramanian et al. 1997], etc. These processes not only increase the thermal budget, but also increase the manufacturing costs. Fabrication on a flexible substrate requires low-temperature processing. Hence, to control the number of GBs in a device, we employ option (b); that is, to scale the device geometry without inducing extra thermal budget and costs (Figure 1).

Because of the similarity of TFT and SOI device structures, we refer to the scaling rules of SOI transistors to form the basis for scaling TFTs. According to the scaling rule for single-crystalline SOI MOSFETs, when the channel length is scaled down, the oxide thickness should also be scaled to improve the gate control over the channel, thereby reducing short-channel effects [Yan et al. 1992]. Scaling of oxide thickness may require fabrication temperature. However, due to the temperature constraint arising from flexible substrates ($<500^{\circ}$C), the oxide thickness in TFTs cannot be scaled as much as that of c-Si SOI MOSFETs corresponding to the same technology node. Therefore, the conventional geometry scaling rule ($>1000^{\circ}$C) is not applicable to TFT device design.

Hence, it is imperative to develop a modified scaling rule under low thermal budget for LTPS TFTs.

It can be seen from the device I-V characteristics that better subthreshold slope (SS) gives lower threshold-voltage (V_t) and higher I_{on} (at iso-I_{off}). A first-order subthreshold-slope equation for SOI transistors can be given as [Suzuki et al. 1996]

$$S = \frac{\ln 10}{\beta}\frac{1}{1-2e^{-\alpha}} = \frac{\ln 10}{\beta}m,$$
$$\text{where } \beta = (KT/q)^{-1}, m = \frac{1}{1-2e^{-\alpha}}, \alpha = \frac{L_{eff}}{2\lambda}, \lambda = \sqrt{\frac{\varepsilon_{si}}{\varepsilon_{ox}}T_{si}T_{ox}}. \quad (1)$$

It can be seen that the subthreshold slope can be improved either by reducing T_{ox} or T_{si} thereby improving the I_{on}-to-I_{off} ratio (assuming that mobility does not vary). Since T_{ox} cannot be scaled in TFTs because of thermal budget, one solution can be to scale the poly-Si film thickness (T_{si}). It is possible to achieve very thin poly-Si films using low-pressure chemical vapor deposition (LPCVD). In Table I, we compare the geometry of c-Si SOI devices with LTPS TFTs at the same technology node ($\sim$200nm). Silicon body thickness (T_{si}) of 35nm is chosen for c-Si SOI because, according to the scaling theory [Yan et al. 1992], this silicon body thickness ($T_{si} \sim$ 35nm) satisfies the performance of a short-channel device (channel length $\sim$200nm) and can be fabricated using available processes. We propose to scale the silicon film thickness without scaling the oxide thickness, to improve the electrical characteristics of a TFT while keeping the fabrication cost low [Li et al. 2007a].

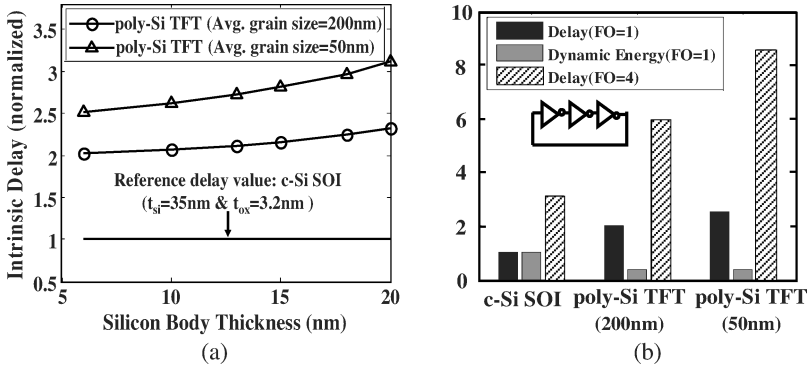

Fig. 3. Delay and dynamic energy of scaled poly-Si TFTs (L ∼ 200nm) in superthreshold region: (a) intrinsic delay with T_{si} scaling in TFTs; (b) delay (FO-1 and FO-4) and dynamic energy in poly-Si TFTs with $T_{si} = 6$nm and $T_{ox} = 10$nm. The values are normalized to and compared with c-Si SOI (L ∼ 200nm).

We analyzed the impact of scaling thickness on the device characteristics in Li et al. [2007a]. Instead of giving an exhaustive device-level analysis, we only discuss the selective power/performance analysis and those results that may be useful to a circuit designer. We designed a three-stage ring oscillator using our proposed TFT and compared it with a contemporary c-Si SOI device at the 200nm technology node. It can be seen that the intrinsic delay (CV/I) of a poly-Si-TFT-based ring oscillator reduces when T_{si} is scaled from 20nm to 6nm (Figure 3(a)). The performance significantly improves as compared to standard TFTs (with $T_{si} = 20$nm), but not as comparable to c-Si SOIs. Also, note that the effective input-gate capacitance seen by an inverter stage is lower in the TFT circuit because of the thicker gate oxide. This reduces intrinsic delay, even with lower drive current. However, in a real circuit, a logic stage drives the interconnect load and several other logic gates. In such cases TFT-based circuits suffer higher delay because of reduced drive current.

Dynamic energy consumption (CV_{dd}^2) is proportional to load capacitance. Since the gate capacitance of a c-Si SOI is approximately three times larger than a TFT (because of thinner gate oxide; Table I), dynamic energy consumption is lower in a TFT-based ring oscillator (Figure 3(b)). Since the delay of a TFT-based ring oscillator ($T_{si} \sim 6$nm) is 2x higher than that of a c-Si SOI ring oscillator (Figure 3 (a)), the V_{dd} in c-Si SOI circuits can be reduced to tradeoff power and performance. With reduced V_{dd} (at iso-performance), we get 33% higher dynamic energy consumption per transition in TFT as compared to that in c-Si SOI. Reduced V_{dd} will also reduce static power consumption in c-Si SOI transistors.

2.3 Manufacturing Perspective: Process Optimization

For a large class of applications, such as portable computing gadgets and medical electronics, very high performance (on the order of GHz) is not necessary, but power consumption is of primary importance. To achieve ultralow power dissipation for such applications, the supply voltage can be scaled below the

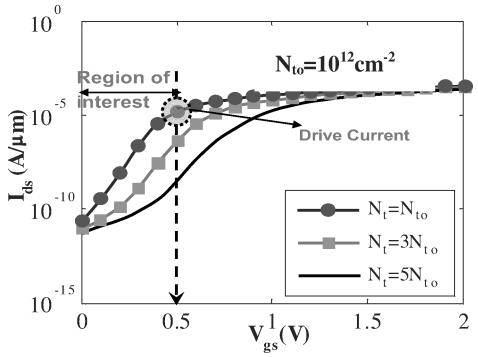

Fig. 4. I_{ds}-V_{gs} curves in scaled poly-Si TFT (L ∼ 200nm, L_{gr} ∼ 200nm) with the variation of mid-gap trap density in subthreshold operation.

threshold voltage such that the device operates in subthreshold regions. So far, subthreshold circuits have been proposed in bulk-silicon technology [Soeleman et al. 1999; Raychowdhury et al. 2005] and it has been indicated that standard device design techniques suitable for superthreshold operation may not be suitable for subthreshold operation. However, by proper process optimization, it is possible to improve the electrical characteristics of TFT devices in subthreshold operation.

In order to evaluate the feasibility of possible device-level optimization for subthreshold operation in LTPS TFTs, first we have to consider the underlying transport mechanisms in TFT, namely drift diffusion in Si grains and thermionic emission across the GB regions. The defects at GBs in poly-Si films consist of "mid-gap states" and "band-tail states" distributing continuously within the forbidden energy-band gap [Chern et al. 1994]. Mid-gap states (due to dangling bonds) and band-tail states (due to lattice mismatch and distortion) have differing impact on the device characteristics of LTPS TFTs. Specifically, mid-gap states mainly affect V_t and SS, while band-tail states strongly influence the field-effect mobility [Chern et al. 1994]. Although the dangling bonds can be reduced by hydrogenation [Chern et al. 1994], lattice distortion cannot be controlled. Since I_{on} exponentially depends on the SS in the subthreshold region, good SS results in higher I_{on} (at iso-I_{off}), as shown in Figure 4. Therefore, apart from their low cost, LTPS TFTs operating in subthreshold regimes can also provide lower power consumption and better performance compared to c-Si SOI devices [Li et al. 2007a].

In order to comprehend the circuit implications, we also evaluated the impact of the proposed device optimization in a three-stage ring oscillator operating in the subthreshold region. Since the effective input-gate capacitance of a transistor consists of a series combination of gate-oxide capacitance and depletion capacitance, the total capacitance in a subthreshold regime is lower than that in a superthreshold regime [Raychowdhury et al. 2005]. Scaled operating voltage and lower input-gate capacitance result in lower intrinsic dynamic energy, albeit with performance loss as compared to superthreshold operation. We also observe that for $T_{si} < 13$nm, poly-Si TFTs show comparable or slightly

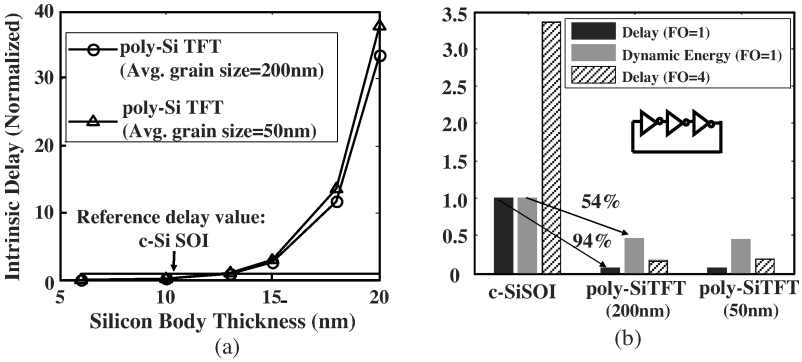

Fig. 5. Delay and dynamic energy of scaled poly-Si TFTs (after proper mid-gap trap-density control) with $L = 200$nm in subthreshold region: (a) intrinsic delay with T_{si} scaling in TFTs; (b) delay (FO-1 and FO-4) and dynamic energy in TFTs (same geometry as Figure 4). The values are normalized to c-Si SOI (same as Figure 4).

better performance than c-Si SOIs in subthreshold operation (with only one grain boundary in the channel and proper control of mid-gap trap density in the grain boundary; Figure 5). Therefore, at iso-performance, the subthreshold poly-Si TFT ring oscillator can achieve much higher savings in dynamic energy consumption compared to the subthreshold c-Si SOI-based ring oscillator (Figure 5). Note that the proceding conclusion is based on the assumption that the trap densities are controlled properly in the subthreshold regime. If this is not true, then the increased leakage current of the LTPS TFT circuit will degrade the I_{on}-to-I_{off} ratio to the point where it is not at all useful. This is further illustrated in Figure 6. The TFT circuits show no benefit in subthreshold operation if the trap density increases beyond 10^{12}cm^{-2}.

2.4 Inherent GB-Induced Variation

Although the nominal performance of a TFT can be significantly improved by device optimization as described before, the statistical variation induced by GBs (Figure 7) can become the dominant factor determining performance compared to other parametric variations (L, W, T_{si}, and T_{ox}). Variation in electrical characteristics can occur because of the variation in number, position, and orientation of GBs [Li et al. 2007a, 2007b, 2007c]. Since GB variation is random in nature, it follows a trend similar to that of random dopant fluctuation in bulk MOSFETs, and becomes worse in scaled technologies (Figure 8). In particular, when TFTs are considered for general circuit applications, the variability in performance can limit the parametric yield of the design. Therefore, it is important to characterize the variation and to design variation-tolerant circuits. In this section, we present a statistical technique to estimate GB-induced variations and evaluate their impact on speed, power, and robustness.

GB-induced variations mainly arise from the statistical GB distribution on a die and can be modeled using a quasi-Manhattan structure (QMS), as shown in Figure 7. In QMS, the grain-size distribution follows the Maxwell distribution [Cheng et al. 2004] along the X and Y directions. Once the circuit layout is

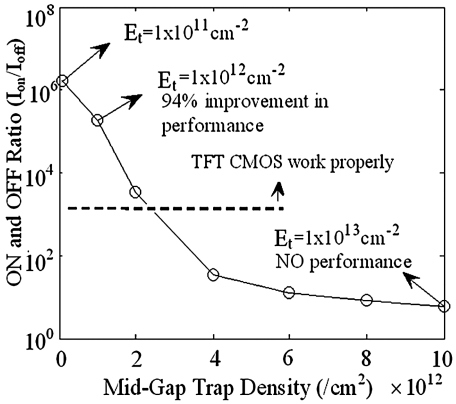

Fig. 6. On- and off-current ratio (I_{on}/I_{off}) of LTPS TFTs with variation of mid-gap-states trap density (10^{11}–10^{13} cm^{-2}).

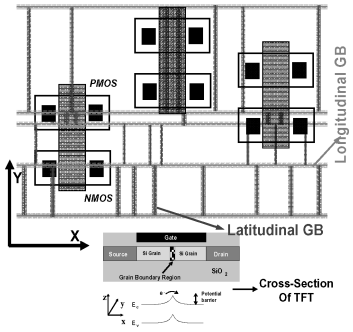

Fig. 7. Process variation due to placement of layouts in polycrystalline film.

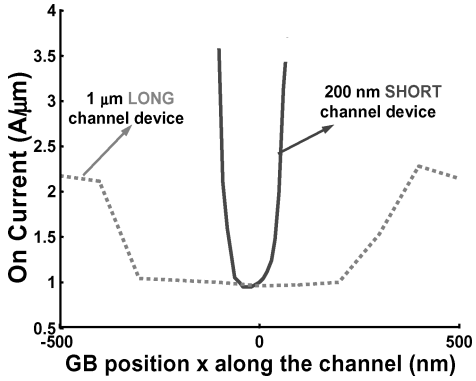

Fig. 8. Position dependency of latitudinal GB in I_{on}. In both I_{on} cases, is normalized to its value when GB is in the middle of the channel.

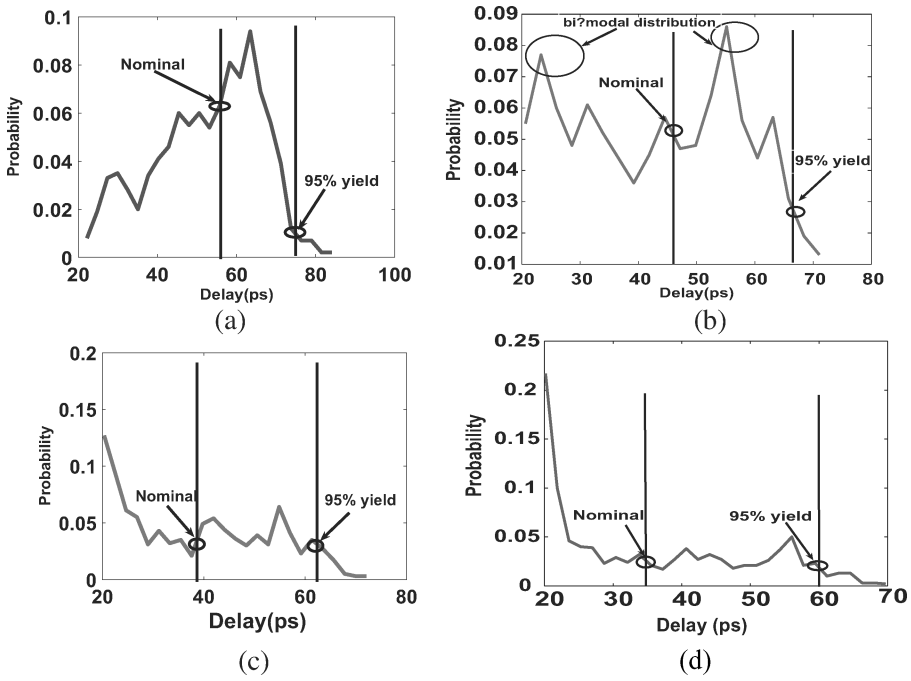

Fig. 9. Statistical delay distribution of an inverter (FO = 1) with average-grain size of: (a) 200nm; (b) 300nm; (c) 400nm; (d) 500nm.

placed on the die, NMOS and PMOS devices in the circuit may have different GB statistics (number, location, and orientation), as illustrated in Figure 7.

Note that the key distribution parameter to model GB distribution on a die is the average-grain size (L_{gr}), which is not constant but process dependent. In this work, we have considered a laser-annealed recrystallization process where the average grain size is typically in the range of 200nm–500nm. Figures 9(a)–(d) illustrate the delay distribution of an inverter (FO = 1) for different average-grain sizes. These figures show that discrete and randomly located GBs produce multimodal delay distributions (contrary to unimodal delay distributions in conventional bulkSis) for average-grain sizes above 300nm. They also show that the overall performance is improved with an increase of average-grain size, but at the cost of large variability (σ/μ). In the following section, we will describe a multifinger (MF) design technique to ensure robust and stable functionality of TFT circuits.

2.5 Multifinger Design

Multifinger designs have been widely used in conventional CMOS in order to reduce the effective capacitance and to improve the layout efficiency [Li et al. 2007b]. In our work, we show that multifingers can actually be used to reduce the effective variation in propagation delay of TFT circuits. In multifinger design, a normal TFT device is divided into multiple subtransistors of the same geometry. The nominal performance (μ) of MF TFTs does not show much

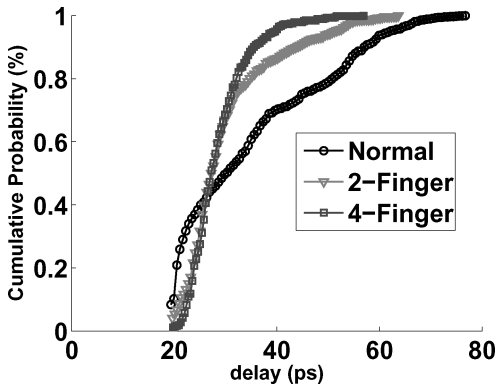

Fig. 10. CDF of propagation delay in an inverter with FO = 1.

difference compared to the original, single, wide TFT structure, since the effective width of the devices is still the same and the performance of the devices is dominated by latitudinal GBs. However, by using multifingers, the variability in drive current of each finger is averaged among the multiple fingers [Li et al. 2007b]. Hence, the MF TFT shows less spread in variation (σ/μ) compared to the single, wide device.

We evaluated the effectiveness of the proposed MF technique on a single-stage inverter with fan-out of 1. Simulation results show that the effective delay variation (i.e., spread of the variation, σ/μ) of a two-finger- and four-finger-TFT-based inverter is reduced by 28.1% and 60.2%, respectively (Figure 10).

For the circuit simulations in the following sections we use the device models for the nominal as well as 95%-yield target case (referred to as the worst, for simplicity) for various average-grain sizes as shown in Figure 9.

3. HYBRID SYSTEM FOR TESTING/VERIFICATION

The proposed LTPS TFTs with high current drivability and less variability have many potential applications (i.e., 3D hybrid system design, etc.). One such application is a TFT-based test engine for reducing testing complexity and improving test confidence [Li et al. 2007]. In the nanometer regime, the continuous scaling of CMOS feature size not only increases fault density, but also introduces new fault mechanisms. To maintain high fault coverage, a large number of test patterns are required, resulting in increased test power, test time, and test cost. Besides offline testing, online testing also plays an important role in maintaining the reliability of nanoscaled systems. An on-chip tester may solve the test problems, but at the cost of design overhead of the testing circuit. Another solution is the addition of design-for-test (DFT) [Abramovici et al. 1995] circuits on-chip. Various DFT circuits have been proposed in the past for online/offline test and calibration [Mukhopadhyay et al. 2005; Ghosh et al. 2006]. One such interesting technique, called CrossCheck [Swan et al. 1989], was proposed to dramatically improve the test and diagnosis capability of the system. However, it comes at the price of die area and test power associated with CrossCheck circuitry. Mysore et al. [2006] proposed a new way to attack

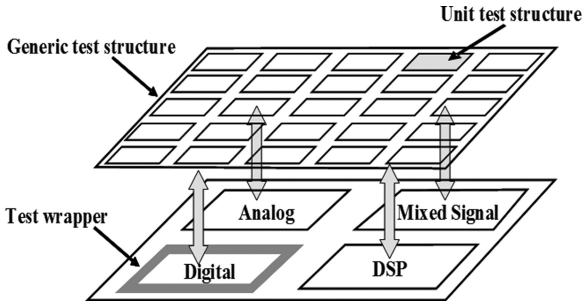

Fig. 11. A conceptual diagram of the generic test structure.

the test problem, with the addition of specialized chip-analysis hardware built on separate active layers stacked vertically on the processor die using 3D Si CMOS technology. This technique shows its effectiveness for helping software developers in tracking down bugs and identifying performance bottlenecks during computer runtime to improve the test capability of the system. However, it limits its application to online monitoring and does not address the issue of offline test. Moreover, the fabrication and integration of the Si CMOS 3D structure is costly. Our test strategy is completely orthogonal to these techniques. We propose a generic and reconfigurable test structure that can be programmed to create any desired logic functionality. To reap the benefits of TFTs, we use it for implementing the test structures. The test architecture is fabricated on a separated layer (i.e., glass) and integrated with the standard Si die, using 3D technology to create a hybrid. The TFT test layer can operate at CMOS-compatible supply voltages with high performance and low power consumption. More importantly, the cost of such TFT-based add-ons is much lower than conventional 3D-chip stacking using CMOS technology. This is because the fabrication process itself is simple and the ease of direct fabrication on transparent substrates significantly reduces the alignment complexity during 3D assembly. The test structure can be configured as an online monitor to track the performance of the processor. It can also be configured as DFT circuitry to reduce the test-circuitry overhead in the Si die for offline test. In this section, we will show the architecture of the TFT-based test structure and the overall test approach. We will also demonstrate TFT application in offline testing and online verification.

3.1 Basic Idea

Figure 11 illustrates the basic concept of the 3D TFT-based reconfigurable test structure for testing systems-on-chip (SoC). The test structure consists of an array of cells called unit test structures (UTSs), each configureable to perform different combinational/sequential logic operations. The UTSs may also contain sensors to measure various circuit parameters (e.g., slew rate, delay, leakage, etc.). Therefore, each UTS may consist of various test resources (a set of logic gates, discrete test components, process/reliability sensors, etc.), a configuration register, and input/output registers; see Figure 12. A UTS can be

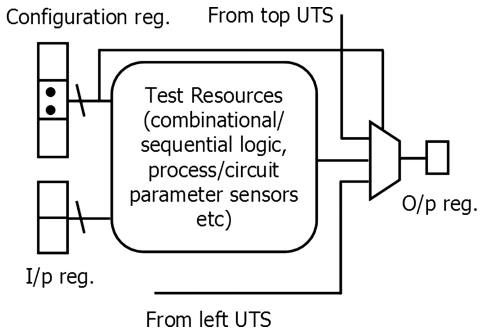

Fig. 12. Block diagram of a unit test structure (UTS).

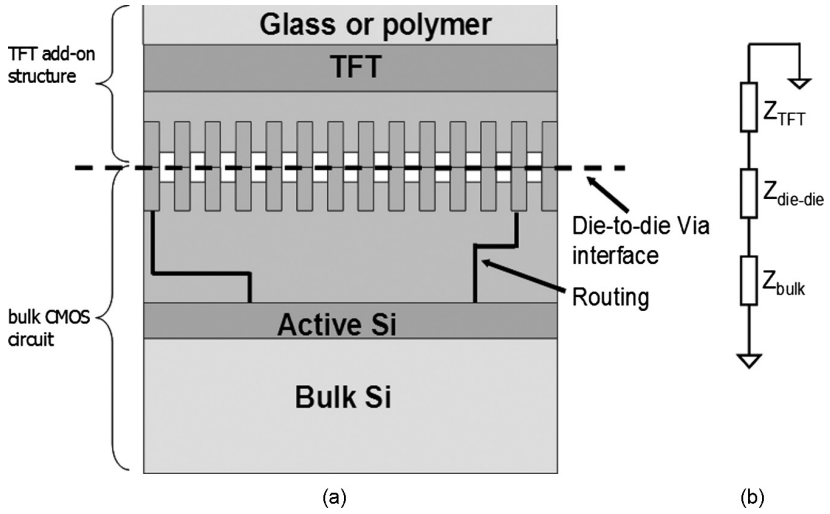

Fig. 13. (a) 3D integration of the TFT test structure with the silicon wafer; (b) model of die-to-die via.

configured to perform a particular function by loading an appropriate pattern in the configuration register. Inputs to the test resources of the UTS can either come from the silicon die or from the input registers, depending on the operating mode of the test structure. During the test mode, the inputs of the UTS are loaded by the tester to the input registers. However, during the normal mode of operation, the inputs come from the node-of-interest on silicon die. The output of the UTS can propagate along the two directions in the 2D plane (in the TFT layer), namely, top-to-bottom (in vertical direction) or left-to-right (in horizontal direction), as shown in Figure 12. The direction of propagation of output can be controlled by providing proper select inputs to the multiplexer from the configuration register.

The TFT test structure can be integrated by face-to-face bonding using die-to-die via interconnects with the silicon wafer, as shown in Figure 13. The die-to-die vias are placed on the top of the metal stack of both dies and are thermally bonded after alignment [Black et al. 2004]. The die-to-die vias on the TFT layer are connected to the inputs and outputs of the UTS, using local

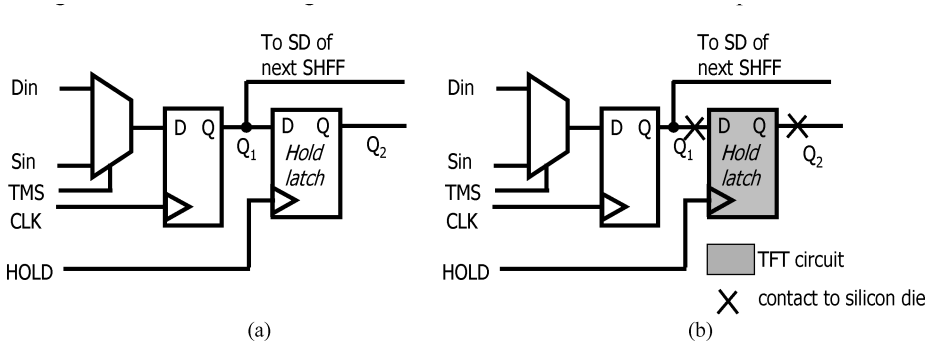

Fig. 14. Normal enhanced scan cell; and (b) hybrid enhanced scan cell with TFT UTS integrated as hold latch.

interconnects and vias. On the other hand, die-to-die vias on silicon can be connected to the node-of-interest through local wires and metal-to-metal vias. Figure 13(b) shows the model of die-to-die vias.

3.2 Test Practice: Offline Test

The UTSs can be configured as a latch, eliminating the need to design the hold latches in silicon and reducing the DFT overhead significantly while having the same flexibility in delay test. In 3D hybrid integration, the inputs/outputs of the UTSs are connected to appropriate nodes of the silicon die through local interconnects and die-to-die vias. Figure 14(a) shows the normal enhanced scan cell, while Figure 14(b) shows the hybrid enhanced scan cell with the TFT UTS configured as a hold latch and integrated with the bulk scan cell. Node "Q1" of the scan cell is connected to the UTS input, and the output of the UTS is connected to node "Q2" in bulk Si. Simulation shows that the delay of the hybrid cell is slightly higher (∼10%) than that of the conventional cell due to the larger capacitance introduced by the die-to-die vias. However, this performance penalty is negligible compared to the considerable area savings (∼15–20%, due to the removal of hold latches from the silicon die [Bhunia et al. 2005]). Furthermore, the power consumption of the hybrid cell is less (∼24% less power) than that of the conventional cell due to reduced gate capacitance ($T_{ox} \sim$ 10nm for TFT and $T_{ox} \sim$ 2nm for bulk), hence reduced power consumption of the TFT hold latch.

Another application of the test structure would be to monitor hard-to-observe nodes during structural or functional test. An entire column of UTSs can be reserved for application as observability points. All hard-to-observe nodes should be connected to these UTSs, which can be configured to act as a buffer. The buffered outputs can either be scanned out to the external tester or compacted, using the UTSs configured as a multiple-input shift register (MISR). In contrast to the 90–95% stuck-at coverage with 2.4K scan test patterns, the proposed structure can provide the same coverage with far fewer patterns, due to increased test points [Carbine et al. 1997].

UTSs can also be configured to build DFT circuitry to assist during testing (manufacturing test or periodic field test). To demonstrate this implementation,

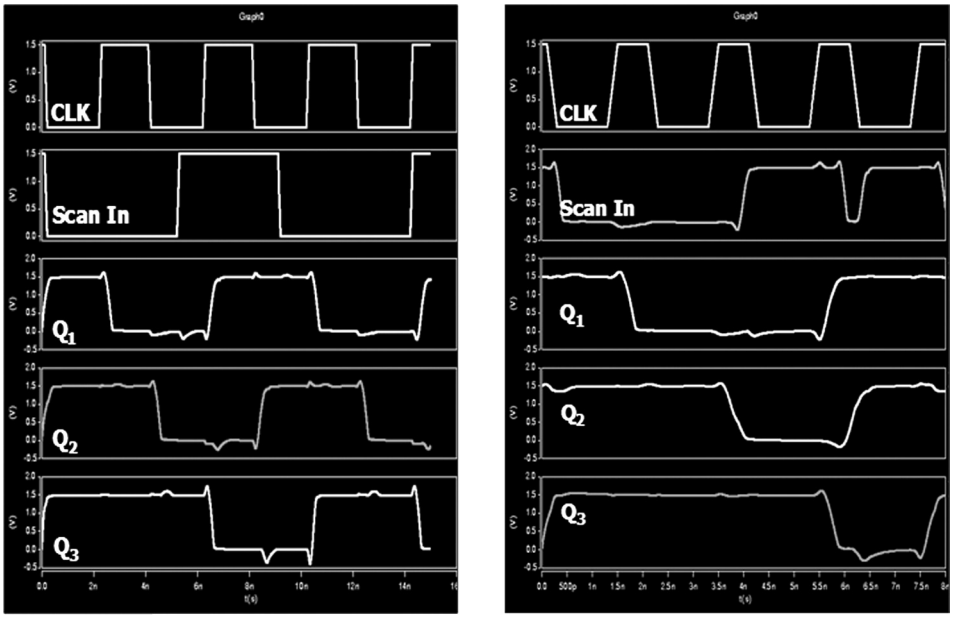

Fig. 15. The simulations results for the scan chain (a) and the LFSR (b) modes.

we have simulated a 3-bit built-in logic-block observer (BILBO) [Abramovici et al. 1995] to operate in all four possible modes (simple latch, scan chain, linear-feedback shift register (LFSR), and multiple-input shift register (MISR)). The simulation results for the scan chain and LFSR modes are shown in Figure 15(a) and (b), respectively. The figure demonstrates the functional correctness of the BILBO, which is designed entirely using TFT devices. We observed that by proper control of grain size and process variation, it may be possible to operate the BILBO at a reasonable speed (0.8–4.3 GHz, compared to 8.4 GHz in conventional bulk). Since the test circuits are reconfigurable, this can be useful in testing a family of processors or DSPs in high-volume manufacturing. Using this technique, the built-in-self-test (BIST) overhead in silicon can be reduced below 5% (LFSR/MISR boundary scan elements/hold latches can be moved to the TFT layer), compared to 15% BIST overhead in conventional design such as the Intel Pentium® Pro processor [Nozuyama et al. 1988].

3.3 Test Practice: Online Test

Besides for offline testing the proposed test architecture can be an effective tool for system designers and software developers for online test and verification of the silicon die. For example, the test structures can also be utilized for online monitoring of certain nodes for diagnosis purposes. Since the clock frequency of the test structure is slower than that of the bulk-Si counterpart (due to slow TFT devices), the test structure can sample the outputs of the test points at low speed. This sampling can either be: (a) compacted, by configuring the

UTS to act as, MISR and the signature can be used later for verification; or (b) provided to the software layer for checking and verification. Since the UTSs can be configured to function as combinational and sequential logic, they also have the potential to be used for repair purposes (at the cost of speed). However, it would require extra effort to connect inputs/outputs of defect-prone logic blocks to the UTS through die-to-die vias. The UTSs can be configured by the software during runtime if repair is required.

Note that the proposed technique is significantly different from that of Mysore et al. [2006] in a number of ways. First the proposed hybrid system is cost effective, since the stacking hardware is only included with developer systems and omitted from consumer systems. Furthermore, 3D stacking with a TFT top layer is much less expensive than a CMOS top layer, while meeting the same functionality. Moreover, direct fabrication on transparent substrate makes LTPS TFTs extremely well suited to 3D integration due to the simple alignment process resulting in further reduced cost. Second, the proposed test architecture can be used for both offline and online test, diagnosis, and possibly self-repair. Third, the test structure is generic and reconfigurable. It can be configured as an online monitor to track the performance of the processor and as DFT circuitry to reduce the test-circuitry overhead in the Si die for offline testing. In a nutshell, our hybrid system not only includes the desirable attributes of the Mysore et al. design [2006], but also has its unique advantages. Further, the proposed hybrid technology is not only limited to testing, it also may fit many other applications due to its compatible performance with CMOS technology and suitability for 3D integration.

4. PRACTICAL CHALLENGES AND ISSUES

In previous sections, we showed the effectiveness of hybrid VLSI systems for testing/online verification and improved reliability. In this section, we address some of the key practical challenges and issues, as follows.

1. *Scalability and Speed/Voltage Mismatch between CMOS and TFT.* Conventional CMOS has been scaled down over generations, making the transistors smaller, faster, and operational at lower supply. Proper device optimization and circuit-level solutions have been utilized to scale down the power supply of conventional TFTs (10–20V) while maintaining sufficient current drivability and less variability. In the proposed 3D hybrid application, the supply voltages of the proposed TFTs have been reduced significantly to make them compatible with bulk CMOS, with reasonable operating speed. However, for proper integration of future scaled silicon technologies (with further reduced supply), the proposed TFT may need to operate in subthreshold regimes (because $V_{t\_bulk} < V_{dd} <= V_{t\_TFT}$). Therefore, it is imperative to devise circuit-level design techniques to allow the TFTs to operate in subthreshold region.
2. *3D Integrations.* Although TFTs provide advantages in terms of 3D integration, there are still issues that need to be further explored and resolved. For example, bonding of the vias, heat dissipation through the 3D structure, etc., are challenges that should be addressed.

3. *Speed Limitations.* The proposed TFTs operate only up to hundreds of MHz. Therefore, they may not be applicable to very high-performance applications; however, they can be used in a large number of DSP applications requiring medium performance.
4. *Design Efforts for Routing.* For different target applications, the designers should route the nodes-of-interest to the die-to-die vias. The task has to be done in close cooperation with the CMOS circuit designers.
5. *Loading of Internal Nodes due to Vias.* The routing of wires that is required to connect the nodes-of-interest to die-to-die vias introduces extra loading to internal nodes. Therefore, such nodes should be chosen carefully so that critical paths are not affected.

5. CONCLUSIONS

We proposed a device/circuit/architecture design paradigm using low-cost LTPS TFTs with CMOS-compatible performance for 3D hybrid systems. Traditionally, TFTs operate at high supply voltages (∼10–20V) due to the presence of highly defective grain boundaries in the channel. To reduce power supply while maintaining sufficient current drivability, we performed device optimizations both at the design phase and manufacturing phase. A scaling rule under low thermal budget (due to flexible substrate) is proposed to keep the number of highly defective grain boundaries (GBs) under control. This will improve performance. We also demonstrate that such an approach can be promising for ultralow-power subthreshold operation, with performance comparable to contemporary single-crystalline SOI devices after process optimization. However, in the optimized device, the GB-induced variation can be a major concern and is more significant than other parametric geometry variations. To ensure robust and stable functionality of the proposed TFT technology, we show that multifinger structures can aggressively reduce the GB-induced inherent variability. Using optimized LTPS TFT technology, we created a generic and reconfigurable hybrid system for application in offline as well as online testing of the underlying complex VLSI circuitry. We believe that such hybrid systems have a promising future in a wide range of new and interesting applications.

REFERENCES

ABRAMOVICI, M., BREUER, M., AND FRIEDMAN, A. 1995. *Digital System Testing and Testable Design.* IEEE Press.

ABRAMOVICI, M., BREUER, M., AND FRIEDMAN, A. 1995. *Digital System Testing and Testable Design* IEEE Press.

BHUNIA, S., MAHMOODI, H., RAYCHOWDHURY, A., AND ROY, K. 2005. A novel low-overhead delay testing technique for arbitrary two-pattern test application. In *Proceedings of the Design, Automation and Test in Europe (DATE),* 1136–1141.

BLACK, B., NELSON, D. W., WEBB, C., AND SAMRA, N. 2004. 3D processing technology and its impact on iA32 microprocessor. In *Proceedings of the International Conference on Computer Design (ICCD)*, 316–318.

CARBINE, A. AND FELTHAM, D. 1997. Pentium (R) Pro processor design for test and debug. In *Proceedings of the International Test Conference (ITC),* 294–303.

CHENG, C. F., JAGAR, S., POON, M. C., KOK, C. W., AND CHAN, M. 2004. A statistical model to predict the performance variation of polysilicon TFTs formed by grain-enhancement technology. *IEEE Trans. Electron. Dev. 51*, 2061–2068.

CHERN, H. N., LEE, C. L., AND FU, T. 1994. Correlation of polysilicon thin-film transistor characteristics to defect states via thermal annealing. *IEEE Trans. Electron. Dev. 41,* 460–462.

GHOSH, S., BHUNIA, S., RAYCHOWDHURY, A., AND ROY, K. 2006. A novel delay fault testing methodology using low overhead built-in delay sensor. *IEEE Trans. Comput. Aided Des. Integr. Circ. Syst. 25*, 2934–2943.

KARAKI, N., NANMOTO, T., EBIHARA, H., UTSUNOMIYA, S., INOUE, S., AND SHIMODA, T. 2005. A flexible 8B asynchronous microprocessor based on low-temperature poly-silicon TFT technology. In *Proceedings of the IEEE ISSCC Tech. Dig.* 272–273.

LI, J., BANSAL, A., AND ROY, K. 2007a. Poly-Si thin film transistors: An efficient and low cost option for digital operation. *IEEE Trans. Electron. Dev. 54*, 2910–2930.

LI, J., KANG, K., AND ROY, K. 2007b. Novel variation-aware circuit Design of scaled LTPS TFT for ultra low power, low-cost applications. In *Proceedings of the International Conference on IC Design and Technology (ICICDT)*.

LI, J., KANG, K., BANSAL, A., AND ROY, K. 2007c. High performance and low power electronics on flexible substrate. In *Proceedings of the Design Automation Conference (DAC*), 274–275.

LI, J., GHOSH, S., AND ROY, K. 2007d. A generic and reconfigurable test paradigm using low-cost integrated poly-Si TFTs. In *Proceedings of the International Test Conference (ITC).*

MUKHOPADHYAY, S., KANG, K., MAHMOODI, H., AND ROY, K. 2005. Design of reliable and self-repairing SRAM in nano-scale technologies using leakage and delay monitoring. In *Proceedings of the International Test Conference (ITC).*

MYSORE, S., AGRAWAL, B., SROVASTAVA, N., LIN, S., BANERJEE, K., AND SHERWOOD, T. 2006. Introspective 3D chips. In *Proceedings of the International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, 264–273.

NOZUYAMA, Y., NISHIMURA, A., AND IWAMURA, J. 1988. Design for testability of a 32-bit microprocessor—the TX1. In *Proceedings of the International Test Conference (ITC)*, 172–182.

PATTI, R. 2006. Three-Dimensional integrated circuits and the future of system-on-chip designs. *Proc. IEEE 94*, 1214–1224.

RAYCHOWDHURY, A., PAUL, B., AND ROY, K. 2005. Computing with sub-threshold leakage: Device/Circuit/Architecture co-design for ultralow-power sub-threshold operation. *IEEE Trans. Very Large Scale Integr. Syst. 13,* 1213–1224.

SOELEMAN, H. AND ROY, K. 1999. Ultra-Low power digital sub-threshold logic circuits. In *Proceedings of the International Symposium on Low Power Electronics and Design (ISLPED).*

STRICKLAND, S., ERGIN, E., KAELI, D., AND ZAVRACKY., P. 1998. VLSI design in the 3rd dimension. In *Integration*. Elsevier, New York, 1–16.

SUBRAMANIAN, V., DANKOSKI, P., DEGERTEKIN, L., KHURI-YAKUB, B. T., AND SARASWAT, K. C. 1997. Controlled two-step solid-phase crystallization for high-performance polysilicon TFT's. *IEEE Electron. Dev. Lett. 18,* 378–381.

SUZUKI, K., TOSAKA, Y., AND SUGII, T. 1996. Analytical threshold voltage model for short channel n+-p+ double-gate SOI MOSFETs. *IEEE Trans. Electron. Dev. 43*, 732–738.

SWAN, G., TRIVEDI, Y., AND WHARTON, D. 1989. Crosscheck—A practical solution for ASIC testability. In *Proceedings of the International Test Conference (ITC),* 903–908.

TANAKA, T., ASUMA, H., OGAWA, K., SHINAGAWA, Y., AND KONISHI, N. 1993. An LCD addressed by a-Si: H TFTs with peripheral poly-Si TFT circuits. *Proc. IEDM Tech. Dig.* 389–392.

WALKER, P. M., MIZUTA, H., UNO, S., FURUTA, Y., AND HASKO, D. G. 2004. Improved off-current and sub-threshold slope in aggressively scaled poly-Si TFTs with a single grain boundary in the channel. *IEEE Trans. Electron. Dev. 51,* 212–219.

WANG, M. AND WONG, M. 2001. Characterization of an individual grain boundary in metal-induced laterally crystallized polycrystalline silicon thin-film devices. *IEEE Trans. Electron. Dev. 48*, 1655–1660.

YAN, R. H., OURMAZD, A., AND LEE, K.F. 1992. Scaling the Si MOSFET: From bulk to SOI to bulk. *IEEE Trans. Electron. Dev. 39,* 1704–1710.

Received August 2007; revised January 2008; accepted April 2008