# Chapter 1
# Variations: Sources and Characterization

**Aditya Bansal and Rahul M. Rao**

**Abstract** This chapter discusses the different sources of variation which can deviate a circuit's characteristics from its intended behavior. First we discuss the sources of process variation during manufacturing, followed by environmental variations during usage. Environmental variations include temperature, voltage fluctuations, and temporal variations. Finally, we discuss the state of art characterization circuits (or sensors) employed to understand the extent and impact of variations.

## 1.1 Introduction

Traditionally, performance benchmarking and power estimation are based on the premise that the electrical characteristics and operating conditions of every device in the design matches the model specifications. However, with continued scaling of device dimensions to sub-100-nm regime, it has become nearly impossible to maintain the same level of manufacturing control and uniformity. This can cause devices to behave differently from model characteristics. Further, devices that were intended to be identical could differ vastly in their electrical characteristics which can lead to functional failures. Environmental factors such as supply voltage and temperature experienced by devices in different parts of the chip (or on different chips) also vary due to different levels of device densities, switching activities, and noise integrity of the various blocks. The voltage and temperature stress experienced by the devices with continued usage degrades their electrical characteristics, thereby increasing the mismatch between the idealistic models and the actual device parameters. All these factors can combine to make the actual design considerably different from the intended design. The performance of the design can thus vary and be lower than the intended one. Similarly, due to the exponential dependence between process/device parameters and transistor leakage, the chip power can also vary and be significantly higher than the nominal values. This translates into a reduced

A. Bansal (✉)
IBM T.J. Watson Research, Yorktown Heights, NY, USA
e-mail: bansal@us.ibm.com

S. Bhunia, S. Mukhopadhyay (eds.), *Low-Power Variation-Tolerant Design in Nanometer Silicon*, DOI 10.1007/978-1-4419-7418-1_1,
© Springer Science+Business Media, LLC 2011

parametric yield, which is the number of manufactured chips that satisfy the required performance, power, and reliability specifications, and hence limits the number of shippable products. With the initial design cost being the same, the cost per good chip increases which has a direct impact on the bottom line dollar value of the design.

The impact of these variations on the parametric yield is generally reduced by guard-banding to provide sufficient tolerance margins during design, which is equivalent to designing at a non-optimal power-performance point. Designs are rated to run at lower than nominal frequencies to guarantee functional correctness in the presence of variations. With increasing variation, the required guard-banding also increases, especially if worst-case conditions are considered. Understanding the various sources of variation and their impact on circuits can help in determining a realistic amount of design margining required, and also provide insights to mitigate the effects of variation. On-chip characterization circuits can help in determining the extent of variation, isolating various effects and their dependence and providing feedback to the manufacturing team. They can also be used in the field to monitor the chip periodically, detect possible failure conditions, and adjust global parameters that can help compensate against these effects.

Variations can be broadly classified into process, environmental, and temporal variations. *Process variations* are a consequence of imperfect control over the fabrication process. *Environmental variations* arise during the operation of the circuit due to changes in the operating conditions such as temperature and supply voltage. *Temporal variations* refer to the change in the device characteristics over time. The rest of the chapter will provide an insight into the sources of variation and circuits commonly used to characterize these effects.

## 1.2 Process Variations

Process variations occur due to lack of perfect control over the fabrication process. These variations always existed since the invention of transistor; however, they are increasingly gaining importance with the scaling of FET dimensions – especially in sub-100-nm technology nodes. In today's VLSI circuits, no two FETs on a chip are exactly same at the atomic level – even neighboring devices can be significantly different. As an illustration, Fig. 1.1 shows the correlation in frequency among four identically designed performance screen ring-oscillators (PSROs) on a chip. It can be seen that the frequencies of neighboring PSROs are correlated to each other; however, not exactly matched though identically designed. This correlation depends on the technology, layout, and the fabrication imperfections.

### *1.2.1 Sources of Variations*

Process variations can be categorized as systematic or statistical in nature. Systematic variations repeat from chip-to-chip and cause fixed offset from design target in FETs. They are typically caused by errors in mask build, optical proximity

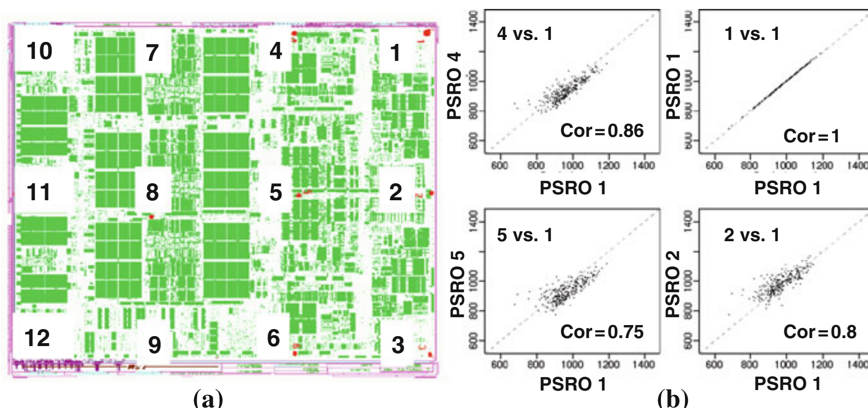



**Fig. 1.1** (**a**) Chip (∼250 mm$^2$) map showing the location of 12 identical ring-oscillators. (**b**) Correlation corner (averaged over 300 chips) for 4 ring-oscillators. (Courtesy: Manjul Bhushan, IBM)

correction (OPC), layout, tool optics, etc. Statistical variations make the two chips or two devices different. Range of statistical variations can vary from ∼300 mm to less than 10 nm. Figure 1.2 summarizes the range of key process variations. Wafer-scale variations (may be due to chuck holding the wafer during process steps) cause variations in wafers but all the chips on a particular wafer see the same effect. Similarly, chip-level variations cause two chips to differ from each other; however, all the FETs

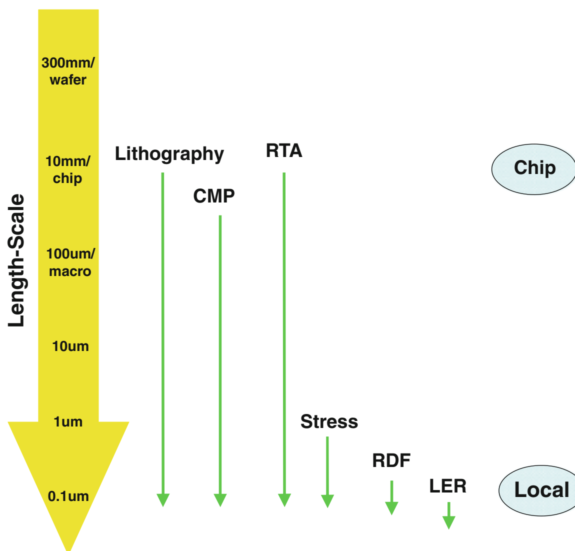



**Fig. 1.2** Variation range of various mechanisms

on one chip see the same affect. Medium range variations (10–100 μm) cause variations from one circuit block to another circuit block. For example, one section of a chip can become very slow whereas another section can be very leaky. Small range variations (< 1 μm) are typically intrinsic and caused by random dopant fluctuations (RDF) and line-edge roughness (LER). These variations cause two neighboring FETs to have different electrical characteristics.

In this section, we will focus on the sources of statistical process variations and their impact on electrical characteristics. We broadly classify sources of variations into two categories – tool-based and intrinsic. Tool-based variations are medium to long-range variations whereas intrinsic variations are small variations.

## *1.2.2 Fabrication Tool-Induced Variations*

### 1.2.2.1 Lithographic Variations

Lithographic variations occur due to the imperfect tools and depend on the physical design. Figure 1.3 shows a simplified view of a lithographic tool used to pattern the FET gate, active diffusion region, contacts, vias, and interconnects. First masks are

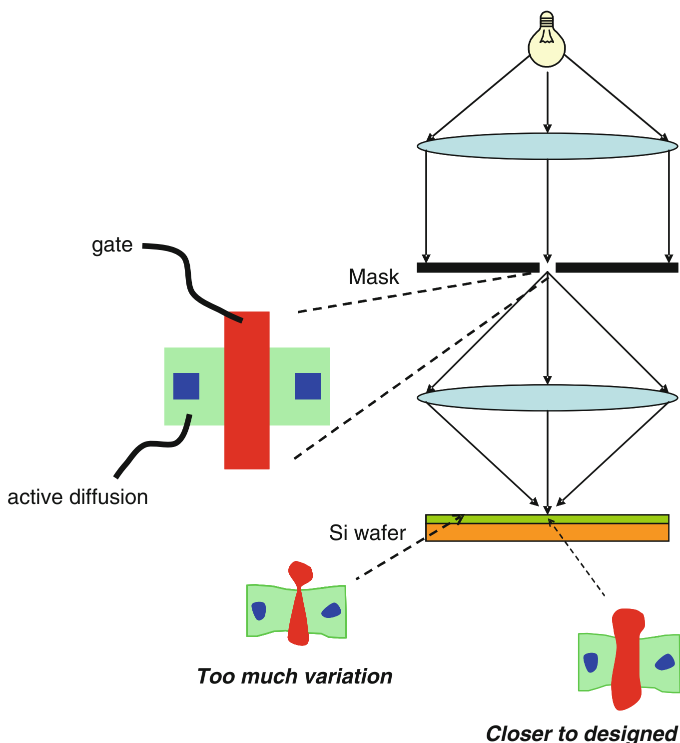



**Fig. 1.3** Schematic of a lithographic tool

generated to represent the designed patterns in each fabrication layer. VLSI circuits are highly complex and the mask formation is largely an automated process. Today, sub-50-nm dimensions are printed on a wafer using a light source with wavelength 193 nm. Several advanced mathematical algorithms and techniques, such as optical proximity correction (OPC) [1], (RET), etc., are used to reliably print a pattern. During the mask generation process, some errors can occur resulting in ill-formation of the pattern. These are called *mask errors*. These errors are random in nature and may or may not exist. For example, mask-error can occur for a specific shape of poly silicon resulting in narrow gate of a FET. After a mask is generated, it is used to print a pattern on the wafer. For sub-50-nm dimensions on a wafer, lenses have to be perfectly aligned and focused. However, it is not possible to perfectly control the focus on the whole wafer resulting in reduced or increased dimensions. For example, *focus variation* in a gate formation can either narrow the gate length or increase it. Further, the dosage of light source is very critical and it is not uniform across the full wafer resulting in the variation in critical dimension (CD) due to *dose variations*. These lithographic variations are typically of the order of tens of microns. These statistical variations can be assumed to have Gaussian distribution. Typically we look at the variation in CD for $\pm 3\sigma$ variations in lithography.

For circuit analysis, an effective gate length needs to be computed for each FET considering process fluctuations. There are several methods devised to account for non-rectangularity in FETs and include them in circuit simulations. Heng et al. [2] proposed one such method which can be used to estimate lithography-aware equivalent gate length and width. This method was used in [3] to design a virtual fabrication environment to maximize circuit yield. Figure 1.4 shows the percentage deviation of gate width and length from the design value in a 45-nm technology as computed using this method. In this example, designed gate length is 40 nm and gate width is 200 nm.

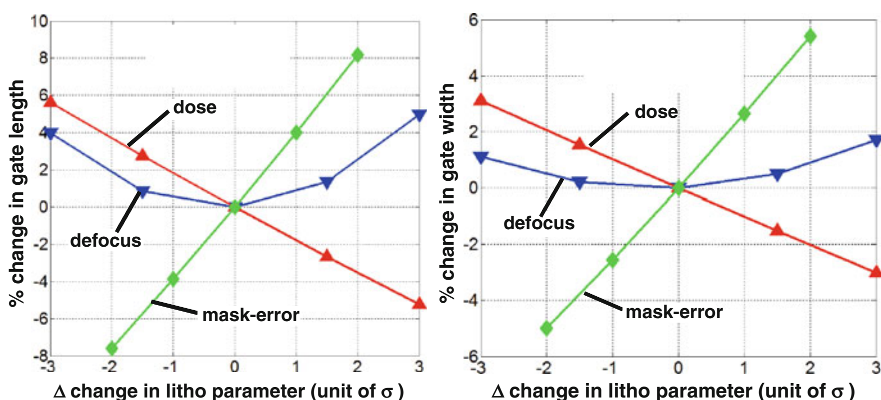



**Fig. 1.4** Variation in a FET's length and width due to variation in lithography parameters – defocus, dose, and mask errors. Percentage change in gate length/width, from designed, are computed using the method proposed by [2]

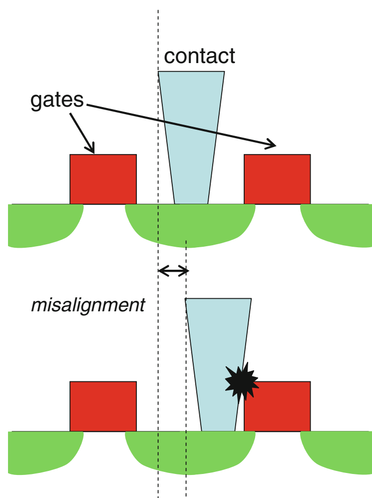



**Fig. 1.5** Implication of contact and gate misalignment

Another source of variation can arise from wafer misalignment from one process to another. One critical example of misalignment-induced variation or defect is the distance between a gate and a contact (Fig. 1.5). For example, gate pitch of 130 nm with gate length of 40 nm leaves 90-nm distance between the two neighboring gate edges for contact formation. For a contact size of 40 nm × 40 nm, the closest distance between the gate edge and the contact edge is 25 nm. The size of the contact cannot be reduced to keep the low resistance while gate pitch cannot be increased to maintain density. Hence, increase in wafer misalignment will increase the gate to contact capacitance eventually leading to catastrophic failure – short between gate and contact.

#### 1.2.2.2 Chemical Mechanical Polishing (CMP)

CMP is used for planarizing the metal interconnect layer or inter-layer dielectric (ILD) between adjacent metal layers due to copper damascene process. Figure 1.6 shows the variation in interconnect thickness in interconnect layer post-CMP. Metal CMP increases resistance due to loss of metal and decreases intra-layer coupling among adjacent wires. Increased resistance is somewhat compensated by decrease in capacitance for circuit delay. ILD CMP introduces change in inter-wire capacitances. Hence, CMP brings variation in the delay of interconnects resulting in non-deterministic circuit performance from chip-to-chip and within chip.

Other fabrication steps which can induce variation are rapid thermal annealing (RTA) [4] and stress liner effect [5]. The key FET parameters prone to these variations are length and width.

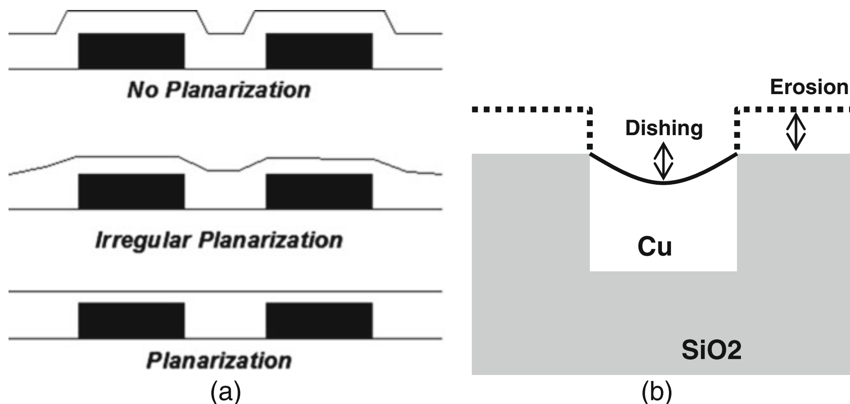



**Fig. 1.6** (**a**) Planarization by CMP and (**b**) dishing and erosion causing variation in interconnect thickness

#### 1.2.2.3 Variation in Contact Resistance

Contact resistance is becoming critical for circuit analysis because of increasing resistance with contact dimension scaling. Figure 1.7 shows the distribution of contact resistance as analyzed by Sleight et al. [6]. They proposed an analytical cone model to compute the resistance under variation in bottom diameter, height and angle of a contact. Variation in contact resistance between two FETs can result in a significant difference in voltages seen by the terminals of these FETs. As an example, let us consider two 1-μm-wide FETs (located spatially far off) such that their source contact resistances differ by 100 Ω. For 1 mA of drain-to-source current,

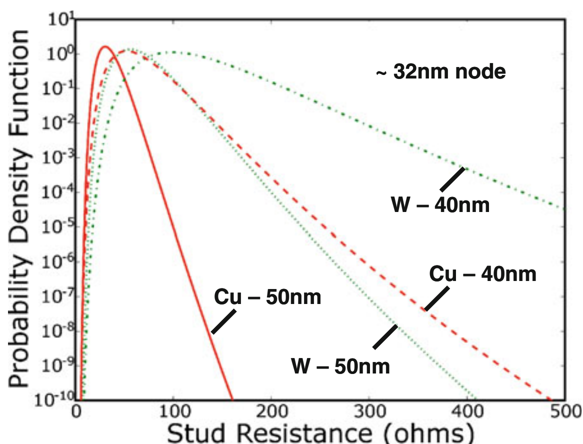



**Fig. 1.7** Distribution of contact resistance due to variation in – angle, bottom diameter, and height for a designed contact diameter. Shown are the two types of contacts – Tungsten (W) and Copper (Cu). W contacts are easy to manufacture. *Source*: [6]

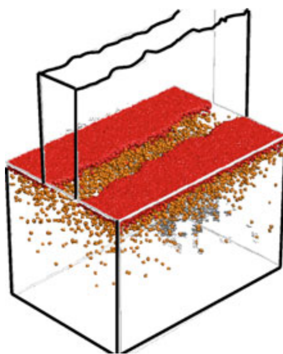
the difference in voltage drops at the source terminals will be 100 mV between the two FETs. This will bring a difference of 100 mV between effective gate-to-source voltages, directly translating to the $V_T$ variation of 100 mV between two FETs. An analytical cone model [6] can be used to compute the contact resistance.

## *1.2.3 Intrinsic Variations*

Beyond variations due to imperfect fabrication tools, some variation sources are intrinsic to a technology. Two key sources of variation which are truly random in nature are random dopant fluctuation (RDF) and line-edge roughness (LER).

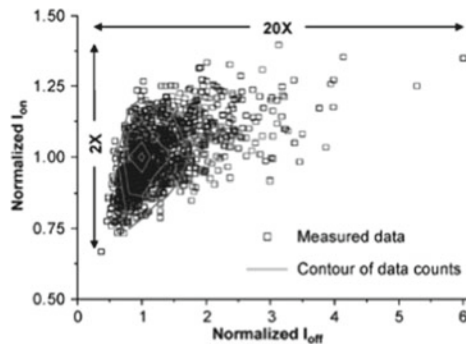
### 1.2.3.1 Random Dopant Fluctuations (RDF)

FETs are doped with impure atoms to control the transistor's electrical properties, primarily short-channel effect and threshold-voltage. With technology scaling, the device volume is reducing thereby reducing the number of dopant atoms in a FET's channel. In sub-100-nm transistors, the position and the number of dopant atoms are critical. When a FET is doped with impurity, the number of dopant atoms and their placements cannot be controlled to the last atom. Moreover, dopant atoms diffuse in the Si channel and randomly get placed as illustrated in Fig. 1.8a. This random number and placement of dopants bring uncertainty in threshold voltage which is called RDF-induced $V_T$ variation. This variation is a grave concern in very small FETs because it is very unlikely to have two neighboring FETs with same number and placement of dopants. The statistical distribution of $V_T$ due to RDF has been found to follow normal distribution. The standard deviation ($\sigma V_T$) of these distribution scales with FET area following inverse square root law. For example, when a FET scales from $L_1$, $W_1$ to $L_2$, $W_2$, the standard deviation changes to



© 2003 IEEE

(a)




© 2009 IEEE

(b)

**Fig. 1.8** (**a**) RDF [7] and (**b**) distribution of Ion and Ioff in a 65 nm technology [8]

$$\sigma_{V_\mathrm{T}}|_{L_2,W_2} = \sqrt{L_1 W_1 \Big/ L_2 W_2}\ \sigma_{V_\mathrm{T}}|_{L_1,W_1} \tag{1.1}$$

implying that if a FET's area is reduced by half, the $\sigma V_\mathrm{T}$ increases by a factor of 1.4 times. Figure 1.8b shows variation in on and off currents due to RDF-induced $V_\mathrm{T}$ fluctuation [8]. As can be understood, large area devices have tighter (lower $\sigma V_\mathrm{T}$) normal distribution of $V_\mathrm{T}$. $\sigma V_\mathrm{T}$ is the same for two FETs with same device widths irrespective of the number of fingers. The distribution of $V_\mathrm{T}$ due to RDF can be obtained by atomistic Monte-Carlo simulations [9, 10].

### *1.2.4 Line-Edge Roughness (LER)*

The gate edge of a MOSFET is not smooth; instead it is rough (as shown in Fig. 1.9a) primarily due to material properties, imperfect lithography and etching tools. Conventionally, polycrystalline silicon (polysilicon) is used to make gate. Polysilicon material has varying grain size, typically in the range of 5–10 nm. Current fabrication tools are unable to smooth this edge resulting in LER. Similar roughness is expected in metal gates due to the properties of metal used and fabrication steps. As shown in Fig. 1.9a, in FETs with gate length 90 nm or higher, the intrinsic fluctuation in gate length due to LER is less prominent compared to 22 nm technologies. In sub-50 nm technologies, the LER can be critical because FETs are typically designed at the knee of $V_\mathrm{T}$ roll-off curve. Figure 1.9b shows that tighter control over the gate length is required in 20 nm long FETs compared to in 65-nm long FETs. The impact of LER can be analyzed by doing 3D TCAD simulations [11]. The distribution of LER-induced $V_\mathrm{T}$ variation is Gaussian and scales with gate width following inverse square root law.

$$\sigma_{V_\mathrm{T}}|_{W_2} = \sqrt{W_1/W_2}\ \sigma_{V_\mathrm{T}}|_{W_1} \tag{1.2}$$

implying that the impact of LER will be less in wider transistors.

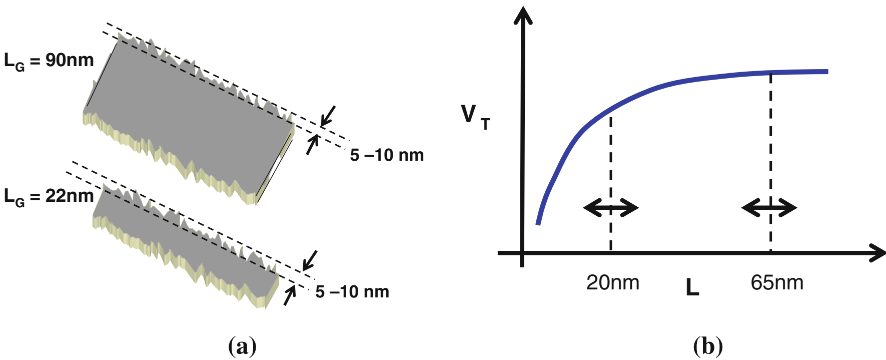



**Fig. 1.9** (**a**) Line Edge Roughness (**b**) $V_\mathrm{T}$ roll-off curve

## 1.3 Temperature and Supply Voltage Variations

The electrical behavior of devices in a circuit is strongly dictated by the operating conditions it experiences. The current through a device is determined by the voltage at its terminals, and hence influenced by the operational supply voltage. The threshold voltage ($V_{TH}$), and carrier mobility ($\mu$) are dependent on the operating temperature. In addition to process variations described in the previous section, any change in these environmental conditions also causes a deviation of the power and performance characteristics of devices from their ideal behaviors.

### *1.3.1 Temperature Variations*

The spatial temperature variation across the die or the temporal difference at the same location on the die depends on several factors. These include:

#### 1.3.1.1 Power Characteristics of the Neighboring Blocks

At a circuit level, the switching activity and the load characteristics of the blocks in the regions around that location determine the power consumption and hence the heat generated near that location. Highly capacitive blocks have a greater current demand and experience a relatively higher temperature in comparison to low-capacitive blocks. Further, blocks with a high-switching activity (e.g., clock-sector buffers) also generate a larger amount of heat in comparison to the blocks with minimal switching activity (e.g., inside a SRAM array or a power-gated block).

#### 1.3.1.2 Thermal Characteristics of Materials

For a given power density, the silicon temperature is a function of the thermal conductivity of the materials used. In a bulk CMOS technology, the heat generated spreads through the silicon substrate as well as the wires. However, in a silicon-on-insulator (SOI) technology the poor thermal conductivity of the buried oxide causes most of the heat to be carried away primarily along the wires, which causes the temperature to increase at a faster rate. This results in a greater variation in the temperature gradient between the power-hungry hotter regions and the low-power cooler regions of the chip.

#### 1.3.1.3 Cooling and Packaging Efficiency

The cooling efficiency of the system also determines the spatial temperature variation. In a conventional design, the rate of temperature increase is determined by the package and heat sink design and the cooling mechanism, with a water-cooled system seeing a better thermal profile than an air-cooled system. However, the temperature variations are likely to be exacerbated in a three-dimensional (3-D) stack technology, where in the dies further away from the heat sink experience a

much reduced cooling efficiency and hence experience a large temperature gradient between the hot-regions and the cold-regions of the chip.

#### 1.3.1.4 Switching Activity and Workload Management

Another factor that determines the spatial temperature variation is the actual workload (or application) being run on the system. Certain blocks are continuously exercised in some workloads. In the current era of multi-core processors, it is likely that some of the cores are inactive and hence cold for a greater percentage of time. Depending on the ability of the system to dynamically manage the workload by periodically assigning tasks to different cores, or moving tasks from one core to another, the temperature difference between the cores will vary.

As a result, the temperature on different parts of the die can be vastly different a given point of time. As an example, Fig. 1.10 shows the temperature profile on an IBM Power 4, with a nearly 20°C difference between the hottest and the coolest portions of the chip. With the power densities of current generation designs being much higher, this spatial temperature variation is likely to be even higher, causing a significant difference in the performances of devices across the chip.

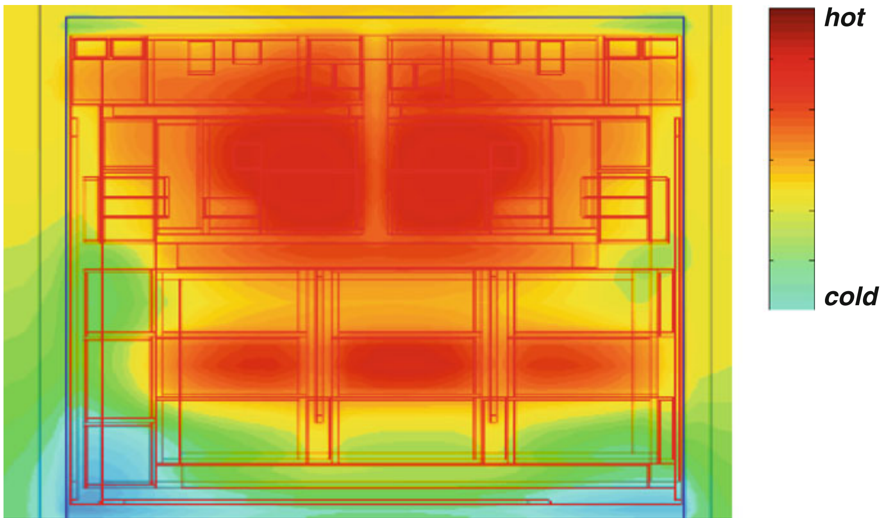



**Fig. 1.10** Temperature difference across regions of IBM Power 4

### *1.3.2 Supply Voltage Variations*

The supply voltage at any location on a die deviates from the nominal design value depending on the characteristics of the power grid network and the current requirement (both steady state and transient) of the design. Increased device

densities and operating frequency of highly pipelined systems have increased the current requirement of current designs. This, along with worsening wire properties, has made the problem of supply voltage variations more significant.

#### 1.3.2.1 Ldi/dt Effect

Various blocks in a design usually share a single power grid. Any sudden increase in the switching activity of a particular block requires additional current to be supplied to that area of the chip. The parasitic inductance in the power grid network and the package causes a *di/dt* drop and hence a transient voltage drop on the supply lines. With aggressive power management techniques, such as clock-gating and power-gating becoming prevalent in designs, the likelihood of such transient current requirements and its magnitude has increased. Further, with multi-core processors under a common power grid becoming the norm, dynamic changes in the current requirement of any one of the cores can create supply voltage variations on all of the others. This can be seen from Fig. 1.11 which shows the droop in the supply voltage for Core 1 and Core 2 in the scenario that Core 2 turns on a few ns after Core 1. In both the scenarios, there is significant supply voltage droop. However, Core 2 has a much worse voltage droop than Core 1, since the voltage droop from Core 1 is coupled to Core 2 just as it turns on, creating an increased supply voltage noise effect. Such peak–voltage droop scenarios are very difficult to predict, model, or avoid and hence require sufficient guard-banding in the design. The impact of such scenarios can be reduced by proper and sufficient use of on-chip decoupling capacitors and regulators.

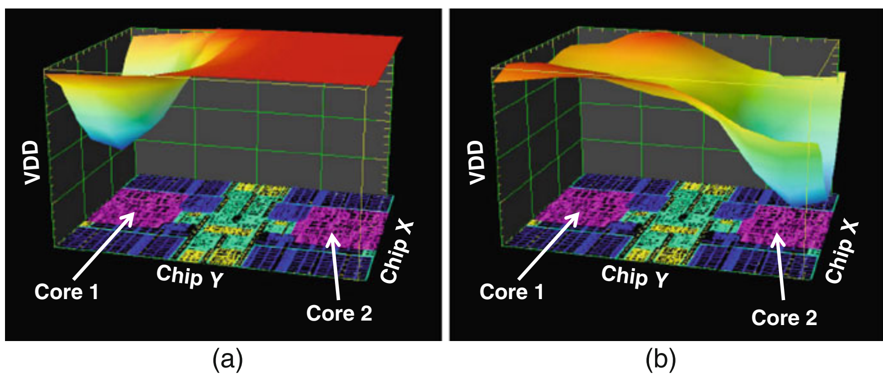



**Fig. 1.11** Power supply droop in a multi-core processor for Core 1 and Core 2. (**a**) Voltage droop in Core 1 (left) as it turns on, followed by (**b**) voltage droop in Core 2 (right) as it turns on a few nano-seconds after Core 1 [12]

#### 1.3.2.2 IR Drop

There is also a voltage drop across the power grid due to the resistance of the power supply lines, which is a function of the steady state current carried through the network. This is referred to as the IR drop and is equivalent to the steady state current drawn from the power supply times the resistance of the power grid. Continued scaling of wire dimensions to match device scaling has increased the resistance per unit length of wire. In addition, growing die sizes with larger transistor counts (especially on-die caches) has increased the steady state (leakage) currents exacerbating the IR drop problem. The resistivity of the wires also increases with temperature, which further worsens the IR drop.

### *1.3.3 Impact on Circuit Performance*

Let us now briefly look into the sensitivity of device characteristics and circuit metrics to these environmental variations.

The mobility of charge carriers degrade with temperature due to the increased scattering in the channel. On the other hand, the semiconductor band gap is known to decrease with temperature. Due to the combination of these contradicting effects, the saturation current (and hence the delay of the worst path in the design) can either improve or worsen. However, in the weak inversion region, the effect of decreased band gap and its exponential influence on the sub-threshold leakage current results in a substantial increase in leakage currents for a small change in temperature. Further, all junction leakages also increase with temperature. Hence, the leakage power significantly increases with temperature making it a critical challenge especially for low-power designs in the hand-held and portable market space. A detailed analysis of the relationship between leakage power and temperature is presented in Chapter 2.

Current characteristics of devices significantly depend on changes in the supply voltage in all regions of operation. With a supply voltage droop of $\Delta V$, both the gate and drain bias seen by the devices reduce, resulting in a reduction in the device current. This increases the path delays and causes non-critical paths to become critical. Further, it worsens signal slew, which can increase the short circuit power drawn and makes the circuits more susceptible to noise.

## 1.4 Temporal Variations

So far we have discussed how a FET's characteristics strongly depend on the fabrication processes and environmental factors. Once a chip is manufactured, packaged, tested (for correct functionality), and shipped to the customers, it is expected to function at the tested voltage, temperature, and frequency till the end of its usage life-time. Assumption is that if we do not have moving parts (as in mechanical machines) then machines do not age (analogous to human aging). However, physical changes can occur in a FET due to movement of charges – electrons and

holes – and breaking of atomic bonds. Key mechanisms which affect a FET's behavior with time are:

(1) *Time-Dependent Dielectric Breakdown (TDDB)*: Creation and joining of defects in gate-dielectric causing gate dielectric breakdown.
(2) *Hot Carrier Injection (HCI)*: Defects in gate stack by highly energized carriers under large lateral (drain-to-source) electric fields causing shift in threshold voltage.
(3) *Bias-Temperature Instability (BTI)*: Capturing of holes (electrons) from the inverted channel in PFETs (NFETs) by the broken Si–H bonds (charge-trapping sites in high-k gate dielectrics such as $HfO_2$).

All the above mechanisms accelerate under large electric fields and high temperatures. In this section we discuss the mechanisms which *vary* the electrical characteristics of a circuit with usage. Many engineers term it as *degradation* of electrical characteristics; however, we will refrain from using the word degradation as in some aspects, electrical aging is not analogous to human or mechanical aging. There are several trade-offs at FET level which pose challenges in meeting desired circuit requirements. Say, if we want to increase performance, we will end up increasing power consumption as well. Therefore, a change in electrical characteristics which will degrade performance as well as power consumption will be a true degradation. We will learn that this is not always true when it comes to FET aging. In this section, we will learn different mechanisms which can affect circuit functionality with usage and also how we can estimate them. Before we proceed, we need to stress that "use" of a FET is a key factor in temporal variation of the characteristics of a FET. When a FET is subjected to voltage bias and operated under certain temperature, materials can degrade resulting in breaking of Si–H bonds causing interface traps, gradual breakdown of gate dielectric due to vertical electrical field, electrons/holes moving in unwanted directions under large lateral, and transversal fields.

### *1.4.1 Bias Temperature Instability (BTI)*

Since the invention of semiconductor transistors, it has been widely known that high voltage and/or temperature stress on the gate electrode with $SiO_2$ as gate dielectric can change the flatband-voltage $V_{fb}$ or the threshold-voltage $V_T$ of MOS transistors [13]. This phenomenon is called Bias-Temperature-Instability (BTI). Several researchers have extensively studied the impact of positive as well as negative high voltage stresses on MOS FETs. With $SiO_2$ as the dielectric, the main focus was only on negative BTI (NBTI) which impacts PFETs. However, with the usage of $HfO_2$ as part of the gate dielectric to facilitate scaling, positive BTI (PBTI) which impacts NFETs has also become significant.

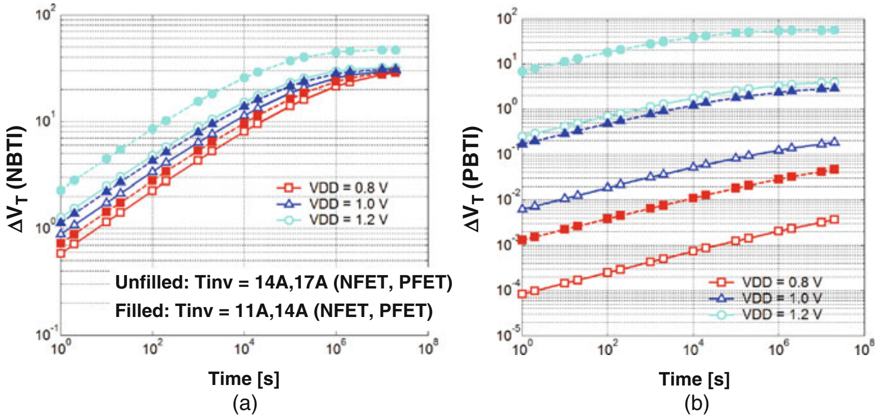



**Fig. 1.12** Time-dependence of threshold voltage in TiN- and Re-gated devices with SiO2/HfO2 as dielectric stack due to (**a**) NBTI (on PFET), and (**b**) PBTI (on NFET). Two cases of inversion thicknesses (Tinv) are shown to analyze the impact of scaling *Source*: [15]

We saw that high temperatures cause $|V_T|$ to reduce, thereby giving a chance to have higher performance; however, mobility also reduces resulting in a conflict and eventually causing higher leakage power and lower performance. Interestingly (and conveniently too), most of the effects of device aging mechanisms can be understood by the change in $V_T$ only. Figure 1.12 shows the time-dependence of threshold voltage shift due to NBTI and PBTI. Scaling the inversion layer thickness ($T_{inv}$) increases the vertical electric field at the given $V_{DD}$ causing increase in $V_T$ shifts due to BTI.

#### 1.4.1.1 Negative-Bias-Temperature-Instability (NBTI)

NBTI relates to the instability induced in a FET due to generation of traps at FET channel and gate dielectric interface. A FET's channel is made of highly ordered crystalline silicon whereas gate dielectric is traditionally made of amorphous $SiO_2$. The interface surface of these two dissimilar materials is rough, resulting in dangling Si atoms from the channel due to unsatisfied chemical bonds. These dangling atoms are the interface traps which can lead to poor performance (lower on-state current) due to charge trapping and scattering. Hence, traditionally, the FETs are hydrogen annealed during fabrication after the formation of Si–$SiO_2$ interface. Hydrogen gas diffuses to the interface resulting in binding of hydrogen atoms (H) to the dangling Si atoms. Figure 1.13 shows the resulting Si–$SiO_2$ interface with hydrogen-passivated Si atoms. These Si–H bonds can break under large vertical negative electrical field ($V_{GS}$, $V_{GD} < 0$) and high temperature. Dissociation of Si–H atoms will result in generation of interface traps. Note that when gate–drain and gate–source voltages are negative in a PFET, channel is inverted and conducting (whereas NFETs are off or under accumulation). Holes in PFET's inversion layer tunnel to $SiO_2$, and are captured by the Si–H bonds thereby weakening the bond.

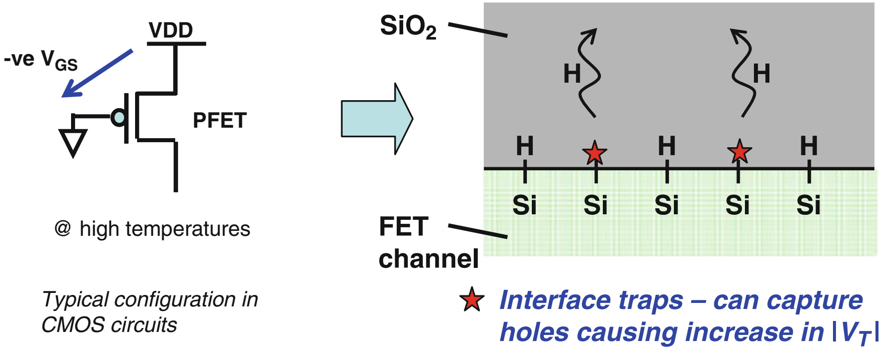



**Fig. 1.13** Schematic of hydrogen dynamics after it is released from a Si–H bond. Interface traps are denoted by * [16]

Subsequently, high temperatures result in dissociation of Si–H pair resulting in one donor-like interface trap or Si dangling bond, and one H atom which diffuses in the oxide. These interface traps can capture the holes in an inverted PFET channel resulting in $V_{GS}$ to become *extra* negative to reach similar strong inversion which existed without interface traps. Hence, $V_T$ of a PFET will become more negative ($|V_T|$ will increase) resulting in reduced on-state current (keeping same $V_{DD}$). Note that such increase of $|V_T|$ in a PFET also reduces off-state leakage power, and effective capacitance thereby reducing active power. This change in $V_T$ of a PFET is called NBTI-induced change in $V_T$.

In early (before the year 2000) technologies, power supply was reduced with gate dielectric scaling to maintain the near constant vertical electrical field. However, in recent technologies, power supply is scaled less aggressively compared to gate–dielectric to improve performance. This has resulted in increased vertical electric field. Further, power density is increasing due to packing of more functionality in same area with every technology resulting in increased local temperatures. Also, oxides are nitrided to prevent Boron penetration from polysilicon gate to the Si channel. Nitrogen is also added in gate dielectric to increase gate dielectric constant thereby increasing gate capacitance and performance. Unfavorably, it increases the NBTI-induced generation of interface traps [17].

Present day VLSI circuits operate at high temperatures (~60–100°C) and also have large vertical electric field resulting in NBTI induced generation of interface traps in PFETs. The interface traps make the $V_T$ of a PFET more negative and reduce carrier mobility. There are several physics models proposed by researchers; the most popular among them is reaction-diffusion model [18–24]. According to R-D model, inversion holes can break Si–H bonds freeing up H atoms (neutral) and creating donor like interface traps. H atoms can either diffuse away from the interface into the oxide or can anneal an existing trap. Interface traps result in increased $V_T$ which can be given as

$$\Delta V_T(t) \approx (m_\mu + 1)\frac{qN_{IT}(t)}{C_{ox}}t^n \tag{1.3}$$

where $m_\mu$ is the mobility-degradation factor, $N_{\rm IT}$ is the number of interface traps at time t, and $C_{\rm ox}$ is oxide capacitance. The time-exponent $n$ is $\sim$ 1/6 [22, 23] for on-the-fly measurements whereas it is >1/6 for delay-based measurements. Here, we will take $n = 1/6$ and include the recovery separately. Along with interface trap generation, other processes such as hole-trapping can also occur [25, 26] changing the form of above equation. Hence, in current technologies, change in $V_{\rm T}$ is empirically estimated by fitting the equation below

$$\Delta V_{\rm T}(t) = A.V_{\rm stress}^{k} {\rm e}^{-Ea/kT} t^{n} \tag{1.4}$$

where parameters $A$, $k$, $n$, and $Ea$ (activation energy) are fitted to match measured data. $V_{\rm stress}$ is the stress voltage. Typically the chip testing time (approximately days–weeks) is significantly shorter than the desired lifetime ($\sim$ 10 years). Hence, accelerated characterization is performed at elevated stress voltages and temperatures, and the models are generated to predict the lifetime of a FET under operating voltage and temperatures.

**Spatial Variation in NBTI**

NBTI primarily depends on the number of interface traps. If a FET size ($L \times W$) is small, the number of traps can differ between FETs resulting in different NBTI degradation even at identical stress conditions. This difference in number of traps between FETs is observed to be random in nature and follow similar scaling law as RDF. Stewart Rausch [27] has given a semi-analytical model to quantify the statistics of $V_{\rm T}$ mismatch between the neighboring FETs in sub-100 nm technologies. The standard deviation of distribution of $\Delta V_{\rm T}$ follows Poisson model (dependence on mean $\Delta V_{\rm T}$ shift) and can be written as

$$\sigma_{\Delta} V_{\rm T} = \sqrt{\frac{2K_1 q T_{\rm ox,eff} {\rm Mean}(\Delta V_{\rm T})}{\varepsilon_{\rm ox} A_{\rm G}}} \tag{1.5}$$

where $T_{\rm ox,eff}$ is the effective oxide thickness of gate dielectric, $\varepsilon_{\rm ox}$ is the dielectric constant of $SiO_2$, $A_{\rm G}$ is the gate area ($L \times W$) and mean ($\Delta V_{\rm T}$) is the mean shift in $V_{\rm T}$ as computed by (Equation 1.4). $K_1$ is an empirical parameter and found to be between 2.5 and 3.0 in most of the measured cases. Inclusion of NBTI statistics for analysis becomes important in small area FETs such as SRAM arrays.

#### 1.4.1.2 Positive-Bias-Temperature-Instability (PBTI)

To reap the benefits of technology scaling, gate dielectric (primarily $SiO_2$) thickness is aggressively scaled to increase gate capacitance, thereby increasing gate control over the channel. In sub-50-nm MOSFETs, due to short-channel effects, $SiO_2$ needs to be scaled to $\sim$ 1 nm. Such thin gate oxides result in high gate direct tunneling leakage currents. Hence, the introduction of high-k gate dielectric became inevitable

in sub-50-nm technologies. $HfO_2$ is one of the most promising candidates as high-k dielectric due to large dielectric constant 20–25, large band gap (5.6 eV), and thermal stability in contact with silicon.

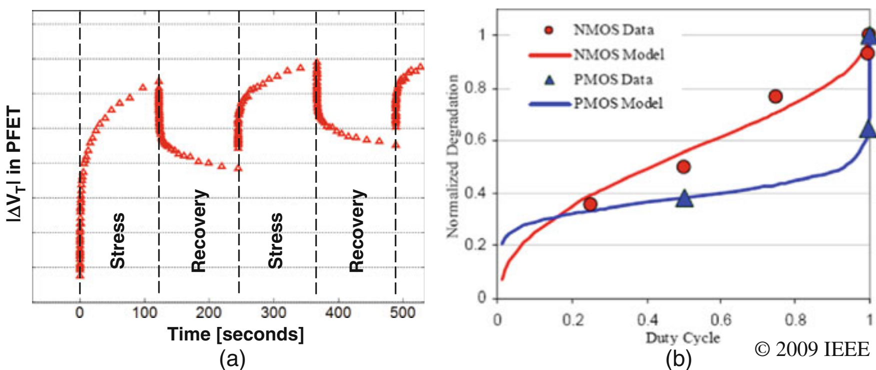
The $V_T$ shift in MOSFETs with high-k dielectric is primarily attributed to charge trapping in high-k layer [28]. The trap density in high-k layer is much higher than $SiO_2$. Hence, at low stress voltages, $V_T$ shifts due to filling of existing traps. Creation of traps is observed at high stress voltages. The reasons for significant charge trapping in high-k gate stacks can be attributed to process-induced defects, impurities, and material-related traps. There have been significant researches in last few years to understand the charge trapping phenomenon [29–32], deemed necessary for optimizing high-k dielectric for VLSI circuits. The $\Delta V_T$ due to PBTI is computed in similar fashion as for NBTI by empirically fitting the Equation (1.4).

#### 1.4.1.3 BTI Recovery

As a FET is turned OFF, it relaxes thereby releasing the trapped charges. Traps can be broadly classified as slow and fast traps. Fast traps get relaxed quickly but slow traps release charges relatively slowly. They can also be classified as recoverable and permanent components of BTI-induced $V_T$ shift [33]. Fraction Remaining (FR), after the stress has been removed, follows the universal recovery curve [34, 35].

$$\mathrm{FR} = \frac{1}{1 + \alpha \left( \frac{T_{\mathrm{relax}}}{T_{\mathrm{stress}}} \right)^n} \tag{1.6}$$

where $T_{\mathrm{relax}}$ is the duration of relaxation and $T_{\mathrm{stress}}$ is the stress time. α is an empirical parameter (fitted separately for NBTI and PBTI) and n is the time exponent used in Equation (1.4). Figure 1.14a illustrates the $\Delta V_T$ with multiple stress-relaxation-stress cycles. Note that as soon as the stress is removed, recovery happens. Figure 1.14b shows the impact of duty cycle on the $V_T$ degradation. Duty



© 2009 IEEE

**Fig. 1.14** (**a**) PMOS Stress-Relax-Stress behavior. (**b**) Alternating OTF measurements for varying duty cycle. Duty cycle of 1 represents the static stress *Source*: [36]

cycle of 1 represents the static stress when the FETs are continuously ON. Duty cycle of 0.5 represents alternating stress causing FETs to be ON 50% of the time. It shows that in a circuit, if we can somehow make sure that a FET does not see static stress, we can increase the life of IC.

Typically we do accelerated stress at very high voltage (sometimes just below the breakdown limit) and then lower the voltage to measure the change in $V_T$. $V_T$ is typically measured using constant drain current method ($V_{GS}$ when $I_{ds}$ is some pre-determined value). FET instantaneously recovers as soon as the gate voltage is lowered to measure $V_T$, making it difficult to measure the actual change in $V_T$. Several measurement techniques have been proposed to circumvent this. One popular method is on-the-fly (OTF) measurement. In OTF, we measure the drain current (let us say with $|V_{GS}| = V_{DD}$ and $|V_{DS}| = 50\,\text{mV}$ at time $t = 0$ and build a relationship $I_{dlin} = k(V_{DD} \cdot |V_T|)^{alpha}$. Now, during measurement, instead of measuring $V_T$, we measure $\Delta I_{dlin}$ and back compute the $\Delta V_T$. Further, recovery varies with the mechanism – NBTI or PBTI – as well as with varying gate stack [37]. Hence, as technology is changed, empirical fitting parameters in BTI-induced $V_T$ shift equations may have to be retuned.

#### 1.4.1.4 Estimation and Impact of BTI in Circuits

In the experiments, FETs are typically stressed using static stress (gate is always ON during stress) and alternating stress (gate is switching at different frequencies with varying duty cycle). These stresses can be conveniently used to generate a threshold voltage model dependent on stress time, frequency and duty cycle. However, in a circuit, more often each FET sees a unique gate signal during the life of usage. This signal depends on

- Applications Executed:

  For example, an application involving multiplication of 40-bit data on a 64-bit multiplier will result in the 24 most significant bits never switching. Another example can be clock signals which always switch irrespective of computation being done in logic blocks.

  Also, in a cache (say SRAM), it is quite possible that some bits are regularly flipped whereas some bits are read often but not flipped resulting in varying impact of BTI between bits. We will study the impact of BTI in SRAMs in more detail in the next section.
- Power Reduction Techniques Used – Dynamic Voltage Scaling (DVS), power–gating, etc.

  In a multi-core chip, all the cores are rarely used at maximum throughput. A core not heavily needed can be run at a lower frequency (by lowering the power supply voltage) to save power. This will result in lower BTI impact in this core. Further, if a core is completely shut down (say using power-gating), it will help in recovery.

There are recent efforts to build commercial tools [38], which can monitor the gate signal of each FET and apply suitable threshold voltage shift using

pre-characterized models. This enables a designer to estimate the EOL shift in a circuit's performance, power, or robustness under real usage conditions. Other efforts are focused to build a BTI aware standard cell library [39].

From circuit application point of view, the nature and analysis of BTI can be broadly classified into two categories – impact of BTI in logic circuits (performance and power are major concerns) and memory circuits (stability and performance are major concerns). In this section, we will focus on these two separately.

### BTI Impact in Logic Circuits

Critical metrics for a logic circuit are performance and power. As we studied earlier, BTI results in increased threshold voltage of NFETs and PFETs which can result in lower currents thereby reducing performance and leakage power dissipation. First let us analyze the impact of BTI on a static CMOS inverter. Figure 1.15 shows the output response of an inverter to an input voltage pulse. The output rise time primarily depends on the strength of PFET to charge the output load to $V_{DD}$, whereas, output fall time depends on the strength of NFET to discharge the output load. NBTI and PBTI will reduce the driving strengths of PFET and NFET, respectively, resulting in increased rise and fall delays. The change in $V_T$ will depend on the duration for which a FET was ON. For example, let us consider two input voltage pulses as shown in Fig. 1.16. In both the cases, the input is high for the duration $t_1$ and low for the duration $t_2$. Let us look at the net FET degradation in both the cases computed by multiplying $\triangle V_T$ due to stress (Equation (1.4)) and fraction remaining (Equation (1.6)):

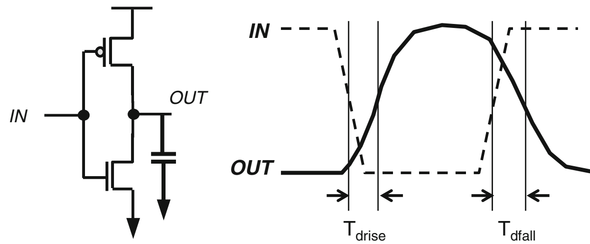



**Fig. 1.15** Schematic and switching waveforms of an inverter

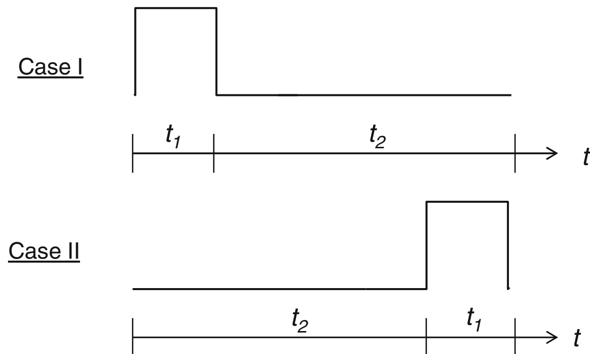



**Fig. 1.16** Two different input signals to an inverter

Case – I:

$$\Delta V_{\mathrm{T}}(\mathrm{PFET}) = F_{\mathrm{NBTI}}(t_2)\ldots(\mathrm{a}) \quad \Delta V_{\mathrm{T}}(\mathrm{NFET}) = \frac{F_{\mathrm{PBTI}}(t_1)}{\left(1+\alpha_{\mathrm{PBTI}}\left(\frac{t_2}{t_1}\right)^{\beta_{\mathrm{PBTI}}}\right)}\ldots(\mathrm{b}) \tag{1.7}$$

Case – II:

$$\Delta V_{\mathrm{T}}(\mathrm{PFET}) = \frac{F_{\mathrm{NBTI}}(t_2)}{\left(1+\alpha_{\mathrm{NBTI}}\left(\frac{t_1}{t_2}\right)^{\beta_{\mathrm{NBTI}}}\right)}\ldots(\mathrm{a}) \quad \Delta V_{\mathrm{T}}(\mathrm{NFET}) = F_{\mathrm{PBTI}}(t_1)\ldots(\mathrm{b}) \tag{1.8}$$

Where $\mathrm{F_{NBTI}}$ and $\mathrm{F_{PBTI}}$ are the $\mathrm{V_T}$ shifts due to stress and computed using Equation (1.4). The FET degradation in the two cases is quite different. Hence, it is important to consider the actual signal at the gate instead of just looking in terms of duty cycle. If instead of looking at actual signal we had assumed that in the total time of $t_1+t_2$, NFET (PFET) is stressed for the duration of $t_1(t_2)$ and relaxed for the duration of $t_2(t_1)$, we would have wrongly estimated the same degradation in both the cases –(Equation (1.7b)) for NFET and (Equation (1.8a)) for PFET. That would have resulted in underestimation of PFET degradation in case I and underestimation of NFET degradation in case II. In practice, during EOL estimation via simulation, it can be quite difficult to monitor the input signal of each FET in a complex circuit. Computation complexity is traded with accuracy by averaging the signal over a long period of time. To date, no such efficient method exists which can accurately estimate the usage-based EOL degradation in each FET of a complex VLSI circuit.

The challenge in front of a circuit designer is to ensure reliable functionality of a circuit during the life of usage. If EOL shift can be accurately predicted during pre-Si circuit tuning, a sufficient guard-banding can be applied such as making extra slack available in critical paths. However, it can be quite difficult to test a circuit for all benchmark applications while monitoring the gate signal of each FET. Hence, on-chip BTI monitors are required to continuously monitor the frequency degradation in a chip and make self-repairs if necessary. The nature of self-repairs can be:

- In multi-core processors, shutting down a core for some time to do BTI recovery
- Power-gating the unused portions of a chip to do BTI recovery along with saving power
- Lowering voltage to reduce stress
- Using redundant copies of the degraded circuits, especially if the primary circuits become unusable due to HCI or TDDB (discussed in next section).

One other important aspect of analyzing BTI impact on a circuit is understanding the relative sensitivity of a circuit's performance to NBTI and PBTI. To a first order, the relative sensitivities of their drive current's strength to $V_{\mathrm{T}}$ can be simplified as

$$\frac{(\partial I_{\mathrm{DS}}/\partial V_{\mathrm{T}})\,|_{\text{pull-down}}}{(\partial I_{\mathrm{DS}}/\partial V_{\mathrm{T}})\,|_{\text{pull-up}}} = \left(\frac{\mu_{\mathrm{electron}}}{\mu_{\mathrm{hole}}}\right)\left(\frac{W_{\text{pull-down}}}{W_{\text{pull-up}}}\right) \tag{1.9}$$

Typically, electron mobility ($\mu_{electron}$) in pull-down NFET is twice of hole mobility ($\mu_{hole}$) in pull-up PFET and width of pull-up ($W_{pull\text{-}up}$) is designed to be higher than the width of pull-down ($W_{pull\text{-}down}$) to balance the rise and fall delays. For example, 50 mV of $V_T$ shift in either PFET (due to NBTI) or NFET (due to PBTI) will result in approximately same amount of increase in RO frequency.

### BTI Impact in Memory Circuits

BTI can be understood as increase in $V_T$ with time. As we saw earlier that the impact of BTI can potentially be different for each FET in a circuit. Hence, the ratioed circuits which primarily depend on the relative driving strengths among different FETs can be severely affected. In this section, we focus on Static Random Access Memory (SRAM) to understand the impact of BTI [15, 40, 41].

Figure 1.17 shows the schematics of an SRAM cell under two conditions – static stress and alternating stress. Static stress happens when a cell is storing the same data for long period of time resulting in a cell becoming asymmetric, i.e., NL (due to PBTI) and PR (due to NBTI) become weak. Note that a cell will undergo static stress even if it is read several times or written with the same data several times. Due to static stress, gradually, the read operation will become unstable in this cell – reading “0” at left-node may cause failure as NL has become weak. We do not consider the impact of PBTI in access FETs (AXL and AXR) because they are turned on for short duration of time. However, in some cases when a cell is accessed very frequently, the access FETs may also become weak resulting in *reduced* impact of PBTI on read stability. Weaker pull-down NFET will also reduce the read current resulting in increased access time. Read current dictates how fast the BL discharges through the series combination of AXL and NL (for $V_L$= “0”). Since reading is getting unstable due to static stress, writing will become easier. For example, flipping the left node from “0” to “1” will become easier because NL has become weak whereas PL retains its original strength. Also flipping the right node from

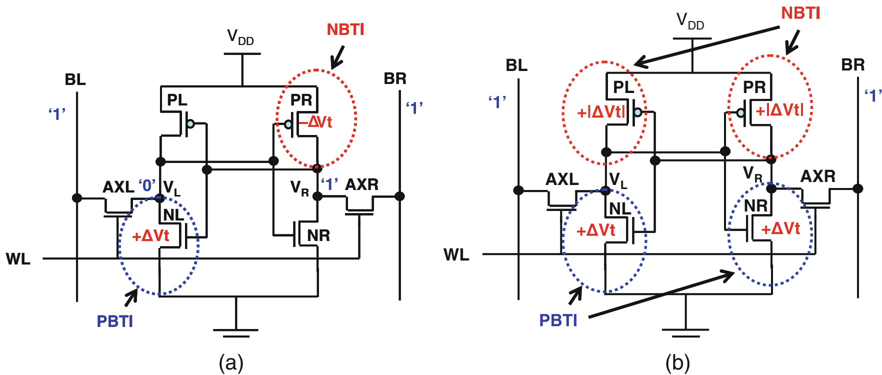



**Fig. 1.17** Schematic of SRAM cell showing the FETs affected under (**a**) Static stress, and (**b**) Alternating stress

"1" to "0" has become easier because PR has become weak whereas NR retains its original strength. Hence, flipping data after long duration of static stress will become easier. On the other hand, a second write to *again flip the data* will become difficult [15].

Alternatively, a cell may be regularly flipped over the period of use resulting in the cell storing "0" and "1" for approximately same duration of time. This will keep the cell symmetric, i.e., same PBTI degradation in NL and NR, and same NBTI degradation in PL and PR (Fig. 1.17b). Considering negligible degradation in access FETs, the read will gradually become unstable and read access time will also increase (due to weak pull-down NFETs). Write operation will become easier, and unlike static stress, even second write operation will also be easier. Further, for a given duration of use, each FET will see the effective stress of half the use time and relaxation for the remaining half. On the other hand, static stress will cause the two FETs (NL and PR) to see the stress during whole use time without any recovery. Note that the static and alternating stress cases are the two extremes. Majority of cells in an SRAM array will see the stress somewhere in between.

In SRAM cells, width of pull-down is typically made higher than the width of pull-up to maintain read stability. Looking at (Equation (1.9)), let us assume that $\mu_{\text{electron}} \sim 2\mu_{\text{hole}}$ and $W_{\text{pull-down}} \sim 2.5W_{\text{pull-up}}$. Now, 50 mV of $V_{\text{T}}$ shift due to PBTI will reduce the drive strength of pull-down NFET roughly $5X$ more than the similar $V_{\text{T}}$ shift due to NBTI will do in PFET. Hence, the stability (read and write) of an SRAM cell is more sensitive to PBTI than NBTI.

### *1.4.2 Hot-Carrier-Injection (HCI)*

Let us consider an NFET (as part of static CMOS inverter) during switching (Fig. 1.18). Its gate voltage rises turning it ON resulting in discharge of output node or drain. As the gate-to-source voltage rises, channel gets inverted resulting in abundance of electrons in the channel. In the initial part of switching, when drain-to-source voltage is high resulting in large lateral electric field, electrons gain high kinetic energy (called hot electrons) and accelerate toward drain. Near drain junction, the hot electrons generate secondary carriers through impact ionization. These primary and secondary electrons can gain enough energy to be injected into the gate stack resulting in the creation of traps at silicon/gate–dielectric interface and also bulk traps in the dielectric. These traps are electrically active and capture carriers at energy levels within the bandgap resulting in increased threshold voltage. HCI occurs in both NFETs and PFETs, however, it is prominent in NFETs because electrons face a smaller potential barrier than holes at the silicon/gate–dielectric interface. This aging mechanism has become less prominent in sub-50-nm technologies due to reduced operating voltages; however, it remains a concern due to high local electric fields. In addition, the dependence of HCI on stress time is higher than BTI bringing it at the forefront of aging culprits after long periods of usage.

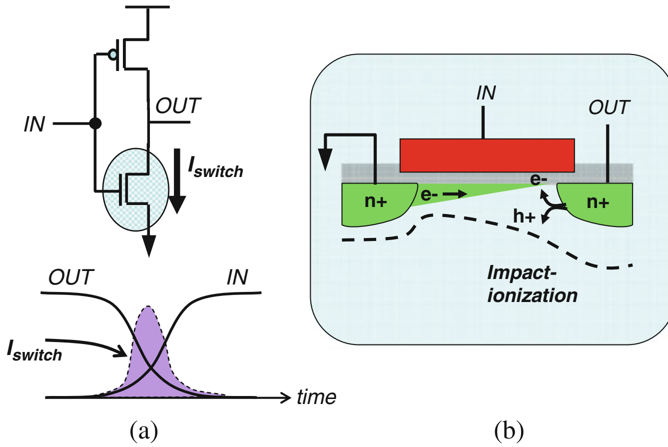



**Fig. 1.18** (**a**) Schematic of an inverter showing switching current during input rising; (**b**) shows the mechanism of charged carriers causing impact ionization

The HCI occurs while a FET is switching contrary to BTI which is dependent on the gate voltage. HCI is empirically estimated in same way as BTI by empirically fitting (equation (1.4)). Authors in [42] measured HCI by separating it from BTI+HCI degradation.

### *1.4.3 Time-Dependent Dielectric Breakdown (TDDB)*

As the name (TDDB) suggests, this phenomena pertains to gate dielectric breakdown under high vertical electric fields. Large voltage drop across the gate stack results in the creation of traps (or defects) within the gate dielectric. These defects may join together and form a conductive path through the gate stack causing dielectric breakdown. TDDB is an increasing concern as gate dielectric thicknesses are scaled down causing *faster* breakdown of gate–dielectric (less defects required to cause breakdown). The physical scaling of gate dielectric has been slowed down with the introduction of high-k dielectric materials; however, it is still a critical aging mechanism.

TDDB is measured by noticing the sudden jump in gate current. However, in today's sub-2-nm gate dielectrics, gate tunneling current is high enough to mask the jump in current. In TDDB analysis, mean-time-to-failure (MTTF) at a particular voltage bias is critical. In low-power circuits, the creation of first conductive or breakdown path through the gate dielectric seldom destroys the transistor. The lifetime of a FET after first breakdown typically depends on the spatial and temporal dependence of the subsequent breakdown [43, 44].

First order impact of TDDB in a circuit can be analyzed by connecting a resistor between gate–source and/or gate–drain in circuit simulations. Instead of a resistor, a constant current source can also be included to simulate the effect of gate current

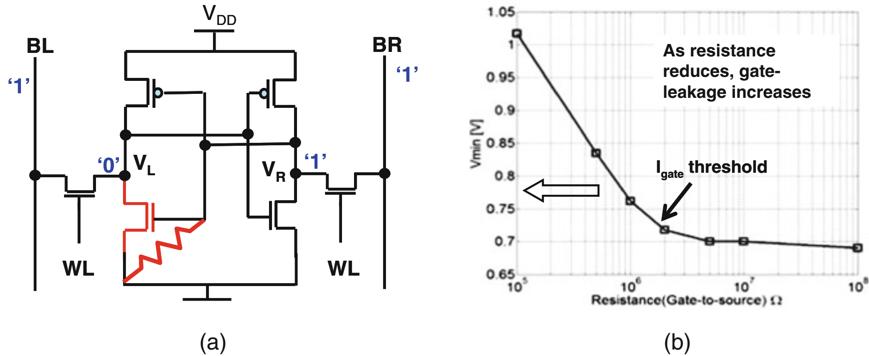



**Fig. 1.19** (**a**) An SRAM cell with breakdown in left pull-down FET simulated by connecting a resistor. (**b**) Change in Vmin with resistance. Igate threshold represents the maximum gate leakage this cell can tolerate before suddent jump in Vmin

during breakdown. For example, Fig. 1.19 shows the simulation of gate–dielectric breakdown in the left pull-down FET of an SRAM cell by connecting a resistor between the gate and source. The value of resistance is varied to obtain the change in minimum operating voltage ($V_{\min}$). This analysis can help a reliability engineer to obtain the gate current ($I_{\text{gate}}$) tolerance of a particular circuit.

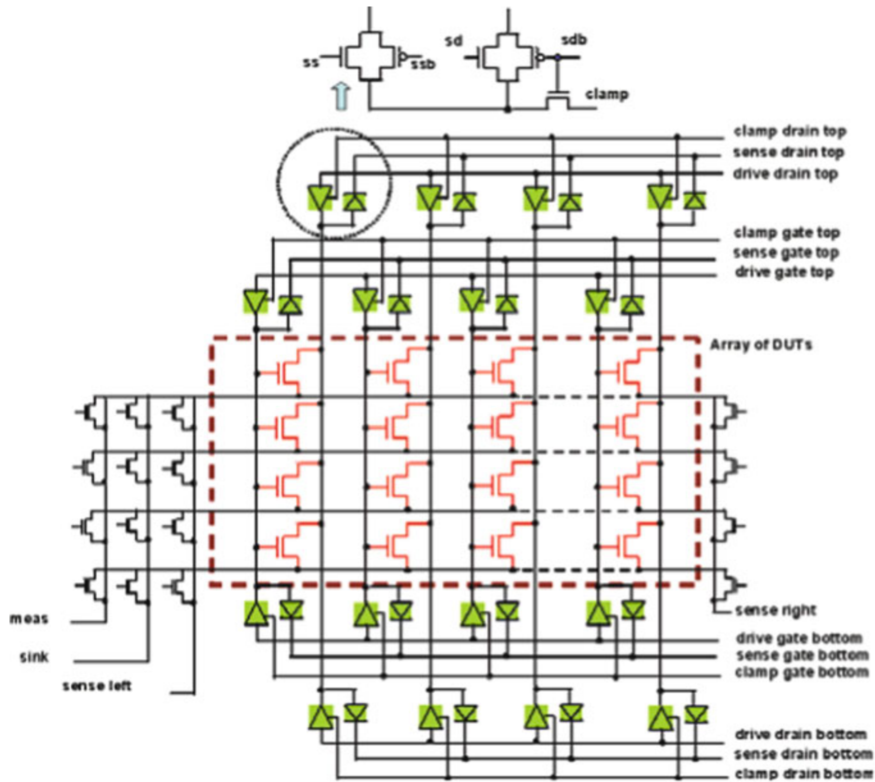
## 1.5 Characterization Circuits

Characterization circuits play a vital role in understanding the sources of variation and their effects. The primary objective of such circuits (or sensors) is to evaluate the extent of variation between the characteristics of a particular device and its ideal expected behavior. In that sense, they play the role of on-chip monitors that are used to assess the health and malaise, if any, of the manufactured design. In earlier technologies, where the manufacturing process was well controlled, it would be sufficient to measure a handful of devices and determine its correctness with respect to the device models. But, with the increase in the extent of variation, the dynamic nature of environmental factors, and the time-dependent behavior of devices, it has become necessary to be able to periodically perform a statistically large number of measurements from monitoring circuits placed at different locations and environments across the die. These measurements help identify and isolate the various sources of variation and their impact on critical circuit parameters such as frequency of operation, noise immunity and power consumption. Further, characterization circuits can be used to determine the amount of guard-banding required during the design phase, and to drive adaptive compensation schemes post-fabrication in an attempt to improve the power and performance characteristics of the design. In this section, we look at several commonly used characterization circuits.

### *1.5.1 Accurate I–V Sensors*

In its simplest form, a characterization circuit consists of a device having each of its terminals connected to probe pads such that its bias voltages can be controlled independently. The current–voltage characteristics of this device can then be measured under different voltage conditions and its electrical parameters such as threshold voltage and source–drain resistances can be estimated and compared with the model specifications. If several such devices can be accessed and measured, a statistical estimation of the extent of variation can be obtained. However, this requires each terminal of each device to be accessible through the use of many probe pads, which makes such an approach prohibitively expensive in terms of silicon area and infeasible.

A multiplexer-based approach can be used to overcome this problem, an example of which is shown in Fig. 1.20. [45]. Here, devices are placed in an $m \times n$ array. To measure the current characteristics of any device, a suitable bias is applied to



© 2006 IEEE

**Fig. 1.20** Accurate I–V measurement sensor [45]. Each of the devices can be individually selected and its current–voltage characteristics measured

that particular row and column, while the rest of the rows and columns are clamped to an unselected state. The current of the selected device under test (DUT) is then measured through the *meas* terminal. To minimize noise effects, the leakage currents of all unselected devices are steered away through a secondary terminal (labeled *sink* in the figure). The impact of undesired voltage drops along the wires and through can be minimized by forcing the gate and drain terminals from both ends of the array. This reduces the IR drop by half and improves supply voltage integrity. In addition, sense nodes are introduced (sense left and sense right) that detect the actual voltage at these sense points, and the applied bias can be adjusted to cancel out any undesired voltage droops. The control of the select signals for the clamp circuits for each column (circled in Fig. 1.20) and the switches for enabling the *meas*, *sink,* and *sense* paths is accomplished through scan chains placed around the array of DUTs.

Devices of various physical dimensions ($L$, $W$), with different number and placement of contacts, physical design arrangements (isolated vs. densely packed), etc., can be seamlessly integrated into the array. This allows for a comprehensive comparison and evaluation of the different variations sources.

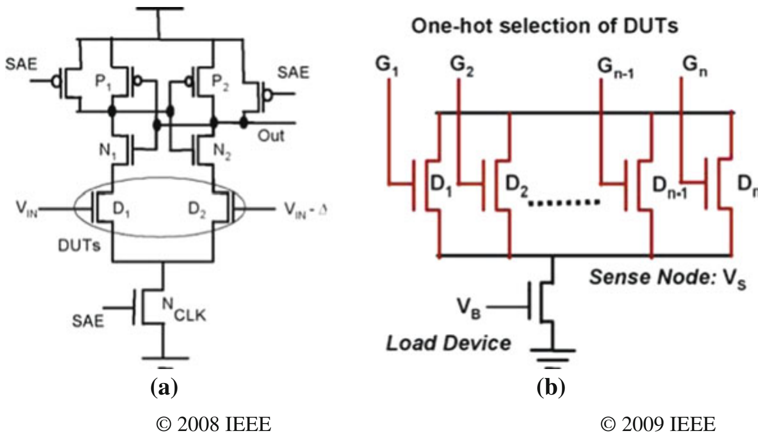
These I–V kind of sensors are very attractive due to their simplicity. The independent control of the terminal voltages allows great flexibility and complete characterization of the devices across the entire range of operation. The complete I–V measurement also gives a high level of confidence in the statistical measurements. However, they rely on accurate current measurements which require precision external current meter terminals (though on-chip current to voltage conversion circuits can be used) and the application of analog voltage at the device (these can also be generated on-chip but add complexity to the circuit). The characterization time is considerably large, since multiple measurements need to be made for each device.

## *1.5.2 Local Variation Sensors*

The objective of local variation sensors is to estimate the mismatch between identical nearby devices without the need for complete I–V measurements. These sensors need to ensure that the effect of any global process or environmental variation is cancelled out. This is usually accomplished by using a differential measurement approach. Local variation sensors can be broadly classified into two sub-categories based on the number of devices that are compared against each other.

### 1.5.2.1 Mismatch Sensors

This type of process sensors detects the mismatch between two identical biased devices using a differential circuit configuration. The sensor is designed such that the output is strongly dependent on the mismatch between the DUTs. During the design of these sensors, it is important to ensure that any mismatch from the non-DUT devices in the differential configuration does not impact the overall measurement accuracy.




© 2008 IEEE © 2009 IEEE

**Fig. 1.21** Sensors for local variation detection (**a**) Mismatch based sensor [46] (**b**) Array based sensors [47]

One example of such a circuit is shown in Fig. 1.21a [46]. A current latch sense amplifier (CLSA) circuit is used as the differential configuration, where $D_1$ and $D_2$ are the DUTs whose mismatch is estimated. The offset voltage for correct sensing of the CLSA depends on the mismatch between the DUTs. This offset voltage is measured by varying the gate bias of one of the two devices ($D_1$ is biased at $V_{\text{in}}$ while $D_2$ is biased at $V_{\text{in}} - \Delta$) and determining the minimum $\Delta$ required for correct transition at the output. The measurement resolution is determined by the granularity of $\Delta$ which can either be applied externally or generated on-chip. A mismatch in the output driver portion of the CLSA can introduce some error into the measurement.

### 1.5.2.2 Array Sensors

Array sensors aim to generate a voltage that is a direct representation of the current (and hence device threshold voltage) in a device selected from an array of devices [48–51]. This sense voltage is then polled for different selections from the array of DUTs, and the distribution of the sense voltage indicate the extent of local random variations.

Figure 1.21b illustrates an example of such a characterization circuit [47]. An array of individually selectable DUTs is arranged in a stacked configuration with a common load device (LD). A first DUT, $D_1$, is selected. The voltage at the sense node $V_s$ depends on the relative strength of $D_1$ and LD. The sense node voltage changes when this $D_1$ is deselected and a second DUT $D_2$ is selected, since it now depends on the strength of $D_2$ and LD. This change in sense node voltage is a representation of the difference between the current characteristics of the two DUTs. All the DUTs can be selected in a cyclical order, and the statistical characteristics of the sense node voltage is computed to determine the extent of local variation.

### 1.5.3 Weak Inversion Sensors

As the name suggests, these sensors consists of DUTs being biased in the sub-threshold region [52, 53]. In the sub-threshold region, the current through the device is an exponential function of threshold voltage, temperature, and supply voltage (due to drain-induced barrier lowering), and hence is extremely sensitive to variation. The principle of operation of these sensors is shown in Fig. 1.22a. The current from a constant current source is fed through a DUT biased in weak inversion, and the voltage at the intermediate node ($V_{out}$) is measured. If the device has a higher threshold voltage than expected, the sub-threshold leakage for a given drain voltage is low. Hence, the drain bias voltage has to increase to cause the leakage current to increase (due to the DIBL effect) and match the reference current. Thus, $V_{out}$ is higher than nominal. Similarly, $V_{out}$ is lower if the DUT has a lower than expected threshold voltage.

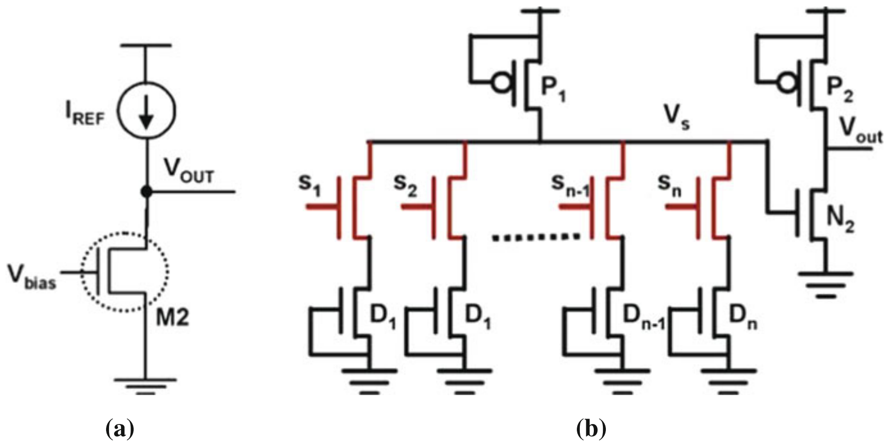



**Fig. 1.22** Weak inversion based sensors (**a**) Principle of operation (**b**) An implementation [53]

An example of a weak-inversion sensor is shown in Fig. 1.22b [53]. Here, an array of DUTs ($D_1$ to $D_n$) is biased in sub-threshold region by grounding their gate terminals. High threshold voltage thick oxide device switches ($S_1$ to $S_n$) are used to select each DUT individually. Let us consider the case where $D1$ is selected, by setting $S_1 =$ "1" and $S_2$ to $S_n = 0$. The voltage at the intermediate node ($V_s$) is then a function of the sub-threshold leakage current of $D_1$ and the load device $P_1$. With the selection of a $D_2$, the sense node voltage changes depending on the mismatch between $D_1$ and $D_2$.

The second stage consisting of $P_2$ and $N_2$ acts as an amplifier and improves the sensitivity of the sensor. This is achieved by sizing $P_1$ such that the sense node voltage $V_s$ biases $N_2$ in sub-threshold region in the absence of variation. A small change in the sense node voltage will result in an exponential change in the current through the second stage, thereby causing a larger change in the output voltage

($V_{out}$). Weak inversion-based sensors are often used as temperature-monitoring circuits, due to the high sensitivity of current to temperature in the sub-threshold or diode-connected configurations.

Another class of sensors are diode sensors, in which the device under test is configured in a diode-configuration by tying the gate and drain of the device (instead of being biased in cut-off region). The readers are encouraged to refer to [54–57] for detailed illustrations.

### *1.5.4 Ring Oscillator Sensors*

Ring oscillator-based sensors are the most popular category of sensors. They characterize variation based on delay through a chain of N gates with the output of the last gate tied back to the input to form a ring. The delay of each gate is a function of the switching characteristics of the devices which varies with the global process parameters and environmental conditions. Since the measured frequency is determined by the sum of $N$ gate delays, these sensors capture global effects and the impact of any local variations is typically averaged out (for sufficiently large value of $N$).

In its simplest form, a ring oscillator-based sensor consists of an odd number of inverter stages (with the theoretical minimum for the number of stages in the ring oscillator being 3) [58, 59], as shown in Fig. 1.23. The frequency of oscillation of the ring is then measured to determine the extent of variation. A higher frequency indicates a fast – leaky design, while a lower frequency of implies a slow, timing critical design. This frequency can be measured externally using a high-precision oscilloscope. Typically, one of the stages of the oscillator is modified to introduce a clamp signal. When the clamp signal is asserted, oscillations are quenched and ring is clamped to a particular state at each of its nodes. This can be implemented by replacing one of the inverter stages in the ring with a Nand2 or Nor2 configuration.

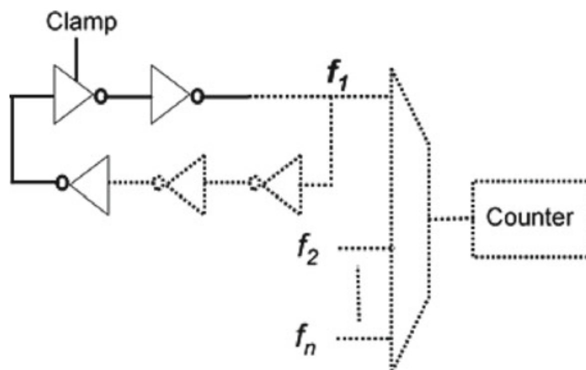



**Fig. 1.23** Ring Oscillator Sensor consists of a chain of gates arranged in a ring. The output of several ring oscillators can be multiplexed and the frequency observed externally or detected on-chip using a counter

Frequency measurement is very attractive since it is independent of any additional delay stages that may be introduced between the output of the ring and the measurement point. This allows for multiplexers to be easily introduced at the output to allow for selection for one of many ring oscillators. Optionally, a counter with a fixed time base can be used to count the number of oscillations of the ring oscillator to get a digital representation of the shift in frequency.

While the Fig. 1.24a shows an inverter as the basic gate, they can be replaced by complex gates. As an example, many ring oscillators can be designed, with each having one of the base cells shown in Fig. 1.24. During characterization, each of the ring oscillators can be selected and its frequency is measured. By studying the spread in frequency of these oscillators, the impact of variations on different classes of circuits can be evaluated. For instance, base cell (b) is made of stacked pull-down trees, and hence is more sensitive to any global shift in NMOS threshold voltages. Similarly, base cell (d) consists of a pass gate in series with an inverter as the base cell and hence can be used to characterize variations in linear mode by suitably biasing the pass gate. In a more generic implementation, each of the stages of the ring oscillator can be a different gate. The only requirement in a ring oscillator-based sensor is that the total number of inversions around the loop be an odd number.

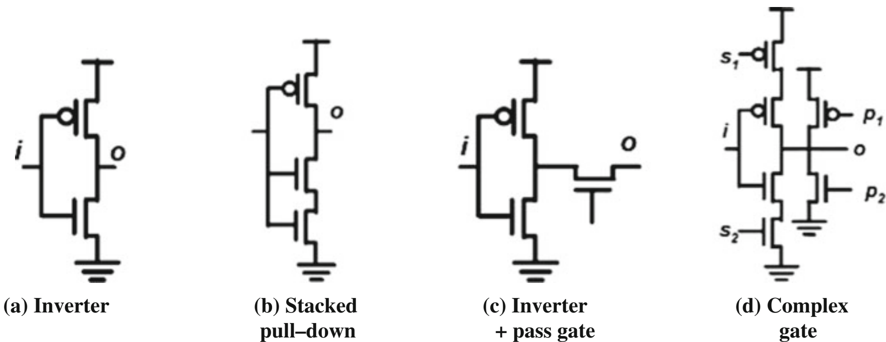



**Fig. 1.24** Different base cells can be used to implement Ring Oscillator Sensors to help isolate the impact of different variation sources (**a**) Inverter (**b**) Stacked pull-down (**c**) Inverter + pass gate (**d**) Complex gate

#### 1.5.4.1 Characterization of Temporal Degradation

Ring oscillator-based sensors can also be used to characterize temporal variations effects [60–62]. By design, the various nodes in the ring oscillator are at logic low and high states for equal duration of time, thereby degrading both NFET and PFET devices. The ring oscillator frequency can be monitored continuously and any degradation in frequency with time can be attributed to device aging affects such as BTI and HCI (described in section 1.4). If device degradation under DC stress condition needs to be characterized, the clamp signal can be asserted to force the ring

oscillator nodes into a particular *stress* state. The ring oscillator then sits in the non-switching state and degrades primarily due to BTI. After any duration of time t, oscillations can be resumed and the frequency is measured. This sequence of *stress* and *measurement* can be repeated at any desired periodicity. The degradation thus measured is a combination of both NBTI and PBTI, and some HCI effects. These effects can be isolated by an intelligent design of the base cell and control circuitry. For instance, the ring oscillator configuration (d) in Fig. 1.24 can be used to isolate NBTI and PBTI. The readers are encouraged to refer to [61] and [62] for a detailed analysis of such configurations.

#### 1.5.4.2 Improving Ring Oscillator Sensitivity by Using On-Chip Reference

While characterizing the impact of temporal variations, it is important to cancel out the effect of any global variations. This can be achieved by using an on-chip reference ring oscillator which is placed in close proximity to the test ring oscillator. The physical proximity of these oscillators ensures that they experience similar global and environmental variations. The reference ring oscillator is not stressed by placing it on a separate supply voltage which is reduced during the stress phase, and hence does not undergo any degradation. In other words, the reference ring oscillator always exhibits its inherent post-fabrication behavior. During measurement, the degraded frequency of the test ring oscillator is compared against the reference ring oscillator. Figure 1.25 illustrates two ways of performing this comparison.

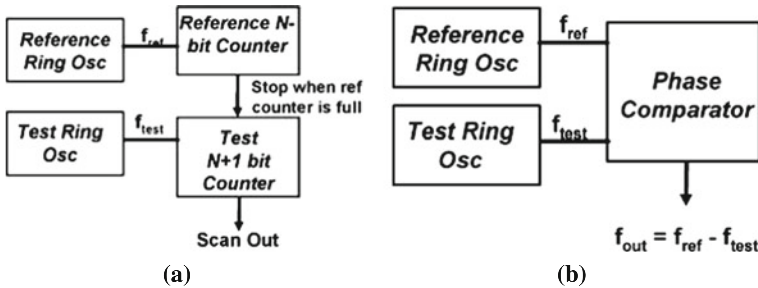



**Fig. 1.25** The sensitivity of ring oscillator based sensors can be improved by using on-chip reference oscillator along with phase comparators on counter circuits

*Counter-Based Comparator*: In a counter-based comparison scheme, the reference ring oscillator sets up a time base that is equivalent to N times its period of oscillation, i.e., $T = 2^N / f_{\text{ref}}$. A second counter, counts the number of oscillations of the test counter during this fixed time duration. When the reference counter hits its maximum value, its stops the count in the test counters. The output of the test counter is now a digital representation of the frequency of the test ring oscillator relative to the reference ring oscillator.

*Phase Comparator*: A phase comparator generates an output frequency which is equal to the difference in the frequencies of the reference and the test

counter. This output signal beats based on the time required for the two oscillators, which are both free running, to align. The beat frequency can then be fed to a counter or measured externally. This differential measurement results in a very high resolution and hence can be used to detect very small amount of degradation in device characteristics.

### *1.5.5 Path Delay Sensors*

Path-delay sensors consist of a chain of gates, with a single transition being launched at the start of this chain. The number of gates traversed by this transition in a given clock period is determined. Path based sensors typically consist of latches and exclusive OR (XOR) gates that act as edge detection circuitry. Fig. 1.26 shows one such implementation. A series of latches can be added to the output of the XOR gates. The sensor is designed such that the transition travels through half the detection chain under nominal conditions in a given clock ($\phi$) period. The location of the edge is detected by comparing the latch output with the expected logic value at that stage of the chain using exclusive OR gates and locating a "1– 0" change in the output. If the global process has shifted toward the fast corner, the delay of each of the gates will be small and the transition will traverse through a larger than nominal number of gates in the given clock period. On the other hand, in the presence of a global shift toward a slow corner, the number of gates traversed will be lower than nominal. Since the launched edge is racing against the clock, it is likely that some of the latches fall into the meta-stability problem resulting in incorrect sampling. A bubble rejection circuitry that detects an isolated "1" or "0" can be used to eliminate that problem.

The resolution of measurement is determined by the gate delay of the elements in the edge-detection circuitry (a fan-out of one inverter delay in this example) and hence is better than ring oscillator-based circuits. The sensing speed is also superior,

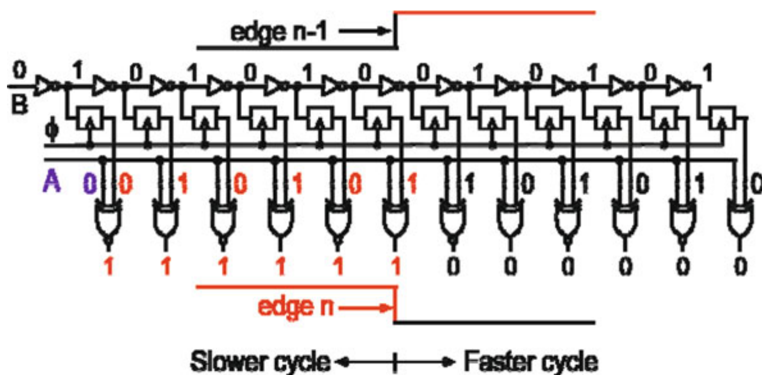



**Fig. 1.26** Path-based sensor consisting of latches and exclusive ORs acting as edge detection circuit [63]

since an input transition can be launched in every cycle. This enables cycle-to-cycle variation to be detected using this type of sensor. For instance, if the activity of a neighboring circuit results in a current surge, and a resultant supply voltage droop, the delay of all the gates in the path will increase, and hence the edge will traverse through a fewer number of gates. By monitoring the output of the edge detector every cycle, the amplitude and duration of the supply voltage droop can be characterized. Path delay-based sensors can also be used to characterize temporal variation effects such as bias temperature instability and hot carrier effects [64]. The type of stress (static vs. alternating) and the duration can be controlled by the signal applied at the input of the chain.

While we have described some commonly used sensors, this is by no means an exhaustive categorization of characterization circuits. In certain scenarios, it becomes necessary to design on-chip sensors that detect more complex circuit parameters. Examples of these are SRAM variation monitors [65], noise margin detectors [66], N-P mismatch detectors [67] etc.

## 1.6 Concluding Remarks

Variations in sub-nanometer technologies have emerged as a critical obstacle to the ability to ship products that meet the system performance, power, and reliability requirements. This trend is expected to worsen in future technologies with newer sources of variation, increased power densities on-chip and bigger chips with larger thermal and supply gradients. Increased thermal profiles and the use of high-k metal gate technologies are also expected to intensify device degradation mechanisms resulting in an increased failure rate in the field. On-chip, characterization circuits that help in understanding these numerous sources of variations, their dependencies, and impact on different circuit styles and parameters are becoming necessary and prevalent in current designs. In addition to proper guard-banding, designs are likely to require intelligent autonomic systems with self-calibration and compensation schemes that help counter the ill-effects of these variation sources and enhance parametric yield.

## References

1. Yu P, Shi SX, Pan DZ (2006) Process variation aware OPC with variational lithography modeling. DAC 785–790
2. Heng FL, Lee J-f, Gupta P (2005) Towards through-process layout quality metrics. Proc SPIE 5756:161–167
3. Bansal et al (2009) Yield estimation of SRAM circuits using virtual SRAM fab. ICCAD 631–636
4. Ahsan I et al (2006) RTA-driven intra-die variations in stage delay, and parametric sensitivities for 65 nm technology. VLSI Tech Sym 170–171
5. Wang C-C et al (2009) Modeling of layout-dependent stress effect in CMOS design. ICCAD 513–520

6. Sleight J.W et al (2006) Challenges and opportunities for high performance 32 nm CMOS technology. IEDM 1–4
7. Hane M et al (2003) Coupled atomistic 3D process/device simulation considering both line-edge roughness and random-discrete-dopant effects. SISPAD 99–102
8. Zhao W et al (2009) Rigorous extraction of process variations for 65-nm CMOS design. TED 196–203
9. Frank DJ et al (1999) Monte carlo modeling of threshold variation due to dopant fluctuations. VLSI Tech Sym 169–170
10. Asenov A, Kaya S, Brown AR (May 2003) Intrinsic parameter fluctuations in decananometer MOSFETs introduced by gate line edge roughness. IEEE TED 1254–1260
11. Asenov A, Jaraiz M, Roy S, Roy G, Adamu-Lema F (2002) Integrated atomistic process and device simulation of decananometre MOSFETs. International conference on simulation of semiconductor processes and devices (SISPAD 2002), Tokyo, pp 87–90
12. Franch R, et al (2008) On chip timing uncertainty on IBM microprocessors. In: Proceedings of the international test conference, Santa Clara, CA, pp 1–7, October 2008
13. Deal BE, Sklar M, Grove AS, Snow EH (1967) Characteristics of the surface-state charge (Q) of thermally oxidized silicon. J Electrochem Soc 114:266–274
14. Shiono N, Yashiro T (1979) Surface state formation during long-term bias-temperature stress aging of thin SiO2-Si interfaces. Jpn J Appl Phys 18:1087–1095
15. Bansal A et al (2009) Impacts of NBTI and PBTI on SRAM static/dynamic noise margins and cell failure probability. J Microelectron Reliab 642–649
16. Küflüoglu H, Alam MA (May 2007) A generalized reaction–diffusion model with explicit H–H2 dynamics for negative-bias temperature-instability (NBTI) degradation. IEEE Trans Electron Devices 1101–1107
17. Kimizuka N, Yamaguchi K, Iniai K, Iizuka T, Liu CT, Keller RC, Horiuchi T (2000) NBTI enhancement by nitrogen incorporation into ultrathin gate oxide for 0.10-pm gate CMOS generation. In: Symposium on VLSI Technology, pp 92–93
18. Jeppson KO, Svensson CM (1977) Negative bias of MOS devices at high electric fields and degradation of MNOS devices. J Appl Phys 48(5):2004–2014
19. Alam MA (2003) A critical examination of the mechanics of dynamic NBTI for PMOSFETs. In: IEDM technical digest, pp 346–349
20. Chen G, Chuah KY, Li MF, Chan DSH, Ang CH, Zheng JZ, Jin Y, Kwong DL (2003) Dynamic NBTI of PMOS transistors and its impact on device lifetime. In: Proceedings of the IEEE international reliability physics symposium, pp 196–202
21. Islam AE, Kufluoglu H, Varghese D, Mahapatra S, Alam MA (2007) Recent issues in negative-bias temperature instability: initial degradation, field dependence of interface trap generation, hole trapping effects, and relaxation. IEEE Tran Electron Devices 54(9)
22. Varghese D, Saha D, Mahapatra S, Ahmed K, Nouri F, Alam MA (2005) On the dispersive versus Arrhenius temperature activation of NBTI time evolution. In: IEDM Technical Digest, Washington, DC, pp 684–687
23. Krishnan A et al (2005) Material dependence of hydrogen diffusion: implications for NBTI degradation. In: IEDM technical digest, 688–691
24. Küflüo˜glu H, Alam MA (2004) A geometrical unification of the theories of NBTI and HCI time-exponents and its implications for ultrascaled planar and surround-gate MOSFETs. In: IEDM technical digest, San Francisco, CA, p 113
25. Mahapatra S, Ahmed K, Varghese D, Islam AE, Gupta G, Madhav L, Saha D, Alam MA (2007) On the physical mechanism of NBTI in silicon oxynitride p-MOSFETs: can difference in insulator processing conditions resolve the interface trap generation versus hole trapping controversy? In: Proceedings of the IEEE international reliability physics symposium
26. Yang T, Shen C, Li MF, Ang CH, Zhu CX, Yeo YC, Samudra G, Rustagi C, Yu MB, Kwong DL (November 2005) Fast DNBTI components in p-MOSFET with SiON dielectric. IEEE Electron Device Lett 26(11):826–828

27. Rauch SE III (December 2002) The statistics of NBTI-induced VT and beta mismatch shifts in pMOSFETs. Trans Dev Mat Rel 89–93
28. Gusev EP et al (2001) Ultrathin high-K gate stacks for advanced CMOS devices. In: IEDM technical digest, Washington, DC, 451–454
29. Zhang JF, Eccleston W (1998) Positive bias temperature instability in MOSFET's. Trans Electron Device 116–124
30. Zafar S, Callegari A, Gusev E, Fischetti MV (2003) Charge trapping related threshold voltage instabilities in high permittivity gate dielectric stacks. J Appl Phys 9298–9303
31. Zafar S (2006) A comparative study of NBTI and PBTI (charge trapping) in SiO2/HfO2 stacks with FUSI, TiN, Re gates. Symposium on VLSI Technology
32. Onishi K et al (June 2003) Bias-temperature instabilities of polysilicon gate HfO2 MOSFETs. Trans Electron Device 1517–1524
33. Grasser T et al (2007) Simultaneous extraction of recoverable and permanent components contributing to bias-temperature instability. IEDM Technical Digest 2007, pp 801–804
34. Rangan S, Mielke N, Yeh ECC (2003) Universal recovery behavior of negative bias temperature instability. In: Proceedings of the IEEE IEDM 2003, pp 341–344
35. Reisinger H, Blank O, Heinrigs W, Mühlhoff A, Gustin W, Schlünder C (2006) Analysis of NBTI degradation- and recovery-behavior based on ultra fast VT-measurements. In: Proceedings of the IEEE international reliability physics symposium (IRPS), Dallas, TX, pp 448–453
36. Ramey S, Prasad C, Agostinelli M, Pae S, Walstra S, Gupta S, Hicks J (2009) Frequency and recovery effects in high-κ BTI degradation. IRPS, pp 1023–1027
37. Denais M (2004) Interface trap generation and hole trapping under NBTI and PBTI in advanced CMOS technology with a 2-nm gate oxide. IEEE Trans Device Mater Reliability 715–722
38. Liu Z, McGaughy BW, Ma JZ (2006) Design tools for reliability analysis. In: Design automation conference, San Francisco, CA, 182–187, July 2006
39. Kumar SV, Kim CH, Sapatnekar SS (2007) NBTI-aware synthesis of digital circuits. In: Design Automation Conference (DAC), San Diego, CA, pp 370–375
40. Kang K et al (2007) Impact of negative-bias temperature instability in nanoscale SRAM array: modeling and analysis. TCAD, IEEE Transactions on Computer Aided Design, pp 1770–1781
41. Lin JC et al (2007) Time dependent Vccmin degradation of SRAM fabricated with high-k gate dielectrics. In: International Reliability Physics Symposium (IRPS), pp 439–444
42. Kim T-h, Wang X, Kim CH (2010) On-chip reliability monitors for measuring circuit degradation. J Microelectron Reliability, 1039–1053
43. Alam MA, Weir BE, Silverman PJ (2002) A study of examination of physical model for direct tunneling current in soft and hard breakdown – part II: principles of area, thickness, and voltage scaling. IEEE Trans Electron Devices 49:239–246
44. Alam MA, Weir BE, Silverman PJ (2002) A study of soft and hard breakdown – part I: analysis of statistical percolation conductance. IEEE Trans Electron Devices 49:232–238
45. Agarwal K, Liu F, McDowell C, Nassif S, Nowka K, Palmer M, Acharyya D, Plusquellic J (2006) A test structure for characterizing local device mismatches. In: Symposium on VLSI Circuits, pp 67–68
46. Mukhopadhyay S, Kim K, Jenkins K, Chuang C, Roy K (2008) An on-chip test structure and digital measurement method for statistical characterization of local random variability in a process. J Solid State Circuits 43(9):1951–1963
47. Rao R, Jenkins K, Kim J (2009) A local random variability detector with complete digital on-chip measurement circuitry. J Solid State Circuits 44(9) 2616–2623
48. Klimach H et al (2004) Characterization of MOS transistor current mismatch. In: Symposium on integrated circuits syst design, 33–38
49. Wang V, Shepard K (2007) On-chip transistor characterization arrays for variability analysis. Electron Lett 43(15):806–806

50. Drego N, Chandrakasan A, Boning D (2007) A test-structure to efficiently study threshold-voltage variation in large MOSFET arrays. In: International symposium on quality electronic design, San Jose, CA, 281–286
51. Terado K, Eimitsu M (2003) A test circuit for measuring MOSFET threshold and voltage mismatch. In: International conference on microelectronics test structures, pp 227–231
52. Kim C et al (2005) Self calibrating circuit design for variation tolerant VLSI systems. In: International online listing symposium, 100–105
53. Meterellyioz M, Song P, Stellari F, Kulkarni J, Roy K (2010) Characterization of random process variations using ultralow-power, high-sensitivity, bias-free sub-threshold process sensor. IEEE Tran Circuits Syst 99:1838–1847
54. Syal A. Lee V, Ivanov A, Altel H (2001) CMOS differential and absolute thermal sensors. In: International online listing symposium, 127–132
55. Chen P, Chen C, Tsai C, Ku W (2005) A time-to-digital-converter-based CMOS smart temperature sensor. J Solid State Circuits 40(8):1642–1648
56. Szekely V, Marta C, Kohari Z, Rencz M (2007) CMOS sensors for on-line thermal monitoring of VLSI circuits. IEEE Trans VLSI Syst 5(3)270–276
57. Chen Q, Meterelliyoz M, Roy K (2006) A CMOS thermal sensor and its application in temperature adaptive design. In: International symposium on quality electronic design, 248–253
58. Zhou B, Khouas A (2005) Measurement of delay mismatch due to process variations by means of modified ring oscillators. Int Sympo Circuits Syst 5:5246–5249
59. Bhushan M, Ketchen M, Polonsky S, Gattiker A, (2006) Ring oscillator based technique for measuring variability statistics. In: International conference on microelectronics text structure, 87–92
60. Karl E, Singh P, Blaauw D, Sylvester D (2008) Compact In-Situ sensors for monitoring negative bias temperature instability effect and oxide degradation. In: International conference on solid state circuits, 410–413
61. Keane J, Wang V, Persaud D, Kim C (2010) An all-In-One silicon odometer for separately monitoring HCI, BTI and TDDB. J Solid State Circuits 45:817–829
62. Kim J, Rao R, Mukhopadhyay S, Chuang C (2008) Ring oscillator circuit structures for measurement of isolated NBTI/PBTI effects. In: International conference on integrated circuit design and Technology, 163–166
63. Drake A. Senger R, Deogun H, Carpenter G, Ghiasi S, Nguyen T, James N, Floyd M, Pokala V (2007) A distributed critical-path timing monitor for a 65 nm high-performance microprocessor. In: International conference on solid state circuits, 398–399
64. Saneyoshi E, Nose K, Mizuno M (2010) A precise tracking NBTI-degradation monitor independent of NBTI recovery effect. In: International conference on solid state circuits, 192–193
65. Bhavnagarwala A, Kosonocky S, Radens C, Chan Y, Stawiasz K, Srinivasan U, Kowalczyk S, Ziegler M (2008) A sub-600-mV, fluctuation tolerant 65-nm CMOS SRAM array with dynamic cell biasing. J Solid State Circuits 43(4):946–955
66. Mojumder N, Mukhopadhyay S, Kim J, Chuang C, Roy K (2010) Self-repairing SRAM using on-chip detection and compensation. IEEE Trans VLSI Syst 75–84
67. Ghosh A, Kim J, Rao R, Chuang C (2008) On-chip process variation detection using slew-rate monitoring circuit. In: International conference on VLSI design, 143–149