

Chapter 2

Variability in Nanometer Technologies and Impact on SRAM

2.1 SRAM Scaling Trends

In today's SoCs and microprocessors, embedded SRAM comprises a large portion of chip area. Figure 2.1 shows an example of a modern microprocessor where embedded SRAM (caches) consumes significant chip area. As shown in Fig. 2.2, SRAM area is expected to exceed 90 % of overall chip area by 2014 [1] because of the demand for higher performance (multiprocessing and multicores), lower power, and higher integration. The large contribution of SRAM has strong impact on chip cost and yield.

To increase memory density, memory bitcells are scaled to reduce their area by 50 % each technology node, as shown in Fig. 2.3. High density SRAM bitcells use the smallest devices in a technology, making SRAM more vulnerable for variations [3, 4]. For example, in state of the art 28 nm technology, a high density bitcell area is approximately $0.12 \mu\text{m}^2$, as shown in Fig. 2.4. This compact bitcell enables an integration of 7.9 Mbit/mm^2 .

While process variation degrades performance and increases leakage in random logic, its impact on SRAM is much stronger. In advanced CMOS technology nodes, the predominant yield loss comes from the increase in process variations, which strongly impacts SRAM functionality as the supply voltage is reduced [5–9]. In particular, local random variations due to and line edge roughness (LER) strongly decrease the robustness of SRAM operation. Figure 2.5 shows that V_{th} variation for SRAM devices increases significantly with scaling, which poses a major challenge for SRAM design [10]. In the following sections we look at the different sources of variations that affect SRAM.

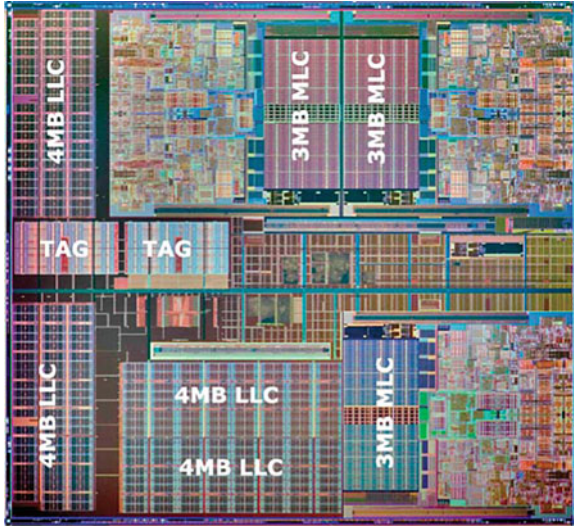


Fig. 2.1 Chip micrograph for a modern microprocessor illustrating the large contribution of embedded SRAM (caches) in the total chip area [2]

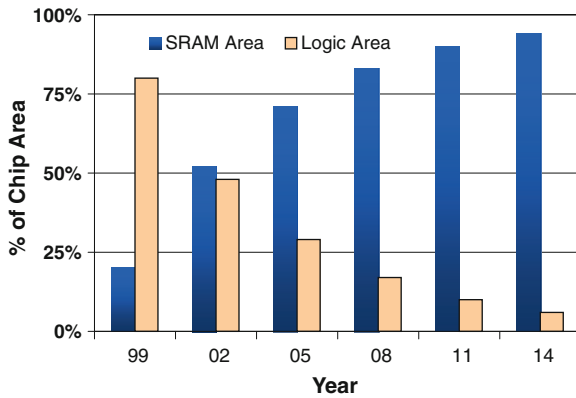


Fig. 2.2 SRAM and logic area versus technology scaling. SRAM dominates chip area in modern SoCs and microprocessors [1]

2.2 Classification of Sources of Variation

Variation is the deviation from intended values for structure or a parameter of concern. The electrical performance of modern IC is subject to different sources of variations that affect both the device (transistor) and the interconnects. For the purposes of circuit design, the sources of variation can broadly be categorized into two classes [12–15]:

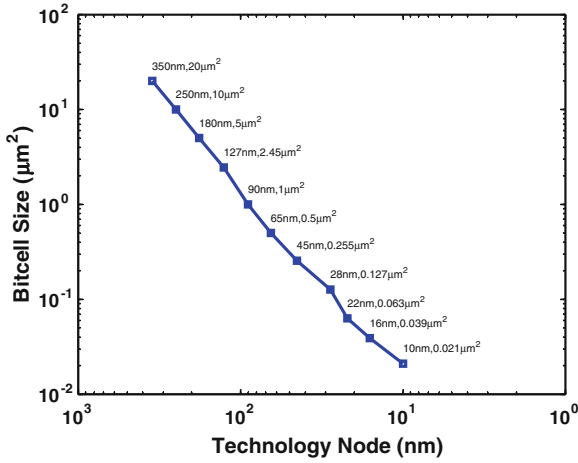


Fig. 2.3 SRAM bitcell area scaling from 350nm down to 10nm technology nodes. Bitcell area continues to scale by 50% for each node

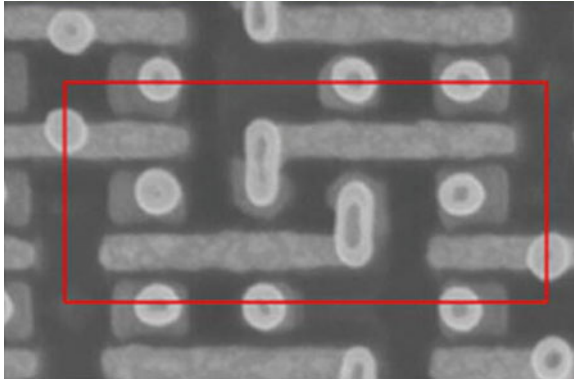
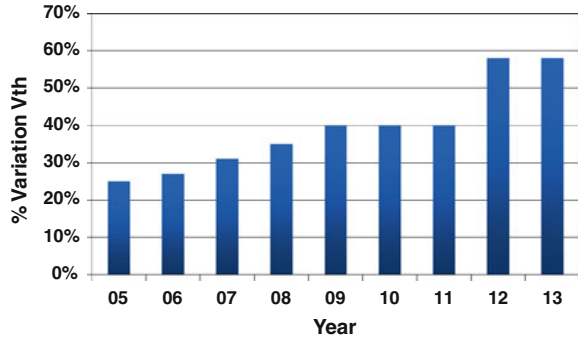


Fig. 2.4 SEM for a $0.12 \mu\text{m}^2$ 6T bitcell in 28 nm technology node [11]

- **Die-to-Die (D2D):** also called global or inter-die variations affect all devices on the same chip in the same way (e.g., they may cause all the transistors' gate lengths to be larger than a nominal value).
- **Within-Die (WID):** also called local or intra-die variations, correspond to variability within a single chip, and may affect different devices differently on the same chip (e.g., devices in close proximity may have different V_{th} than the rest of the devices).

D2D variations have been a longstanding design issue, and are typically accounted for during circuit design with using corner models [12, 13, 16]. These corners are chosen to account for the circuit behavior under with the worst possible variation, and

Fig. 2.5 SRAM devices V_{th} variation scaling trend [10]



were considered efficient in older technologies where the major sources of variation were D2D variations.

However, in nanometer technologies, WID variations have become significant and can no longer be ignored [17–23]. As a result, process corners-based design methodologies, where verification is performed at a small number of design corners, are currently insufficient.

WID variations can be subdivided into two classes [12–15]:

- **Random variations:** as the name implies, are sources that show random behavior, and can be characterized using their statistical distribution.
- **Systematic variations:** show certain variational trends across a chip and are caused by physical phenomena during manufacturing such as distortions in lenses and other elements of lithographic systems. Due to difficulties in modeling this type of variation, they are usually modeled as random variations with certain value of spatial correlation.

Other classifications for variability include time-dependency (long or short), static, dynamic, device, interconnect, and environment. In the following sections, we present an overview of the device, interconnect, and environment sources of variations.

2.3 Device Variability

Process variations impact device structure and therefore change the electrical properties of the circuit. In the following subsections, we review the main sources of variations that affect device performance.

2.3.1 Random Dopant Fluctuations

As CMOS devices are scaled down, the number of dopant atoms in the depletion region decreases, especially for a minimum geometry device. Due to the discreteness of atoms, there is statistical random fluctuation of the number of dopants within a

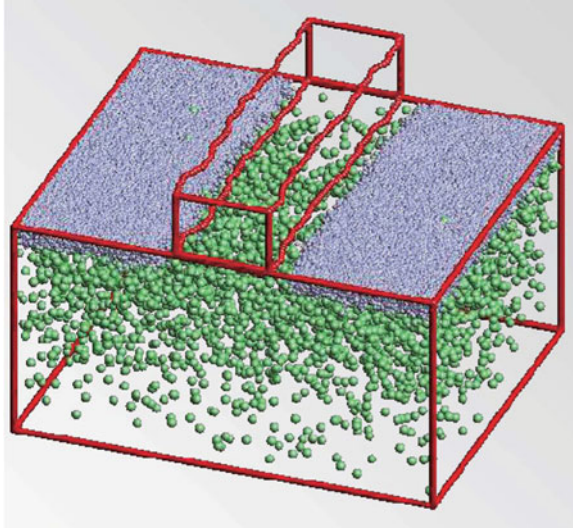


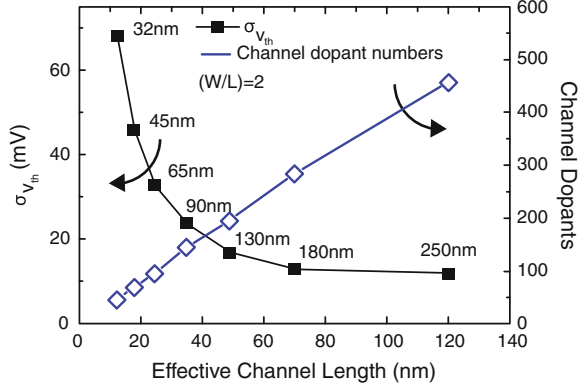
Fig. 2.6 Atomistic process simulation incorporating random dopant fluctuation (RDF) and line edge roughness (LER) as the sources of intrinsic fluctuation [30]. The *dots* show the dopant atoms that determine the device's threshold voltage.

given volume [24–29]. This fluctuation in the number of dopants in the transistor channel results in variations in the observed threshold voltage V_{th} for the device. Figure 2.6 shows how dopants are placed in the transistor channel.

For example, in a uniformly doped $W = L = 0.1 \mu\text{m}$ NMOS, if the doping concentration is $N_a = 10^{18} \text{cm}^{-3}$ and depletion width at zero body bias is $W_{dmo} = 350 \text{\AA}$, the average number of acceptor atoms in the depletion region can be calculated as $N = N_a \cdot L \cdot W_{dmo} = 350$ atoms. Due to the statistical nature of dopants, the actual number fluctuates from device to device with a standard deviation following a Poisson's distribution, and therefore $\sigma_N = \langle (\Delta N)^2 \rangle^{1/2} = \sqrt{N}$, which for our example yields $\sigma_N = 18.7$, a significant fraction of the average number N (σ_N/N is 5% in this example). Variation in the number of dopant atoms directly affects the threshold voltage of a MOSFET, since V_{th} depends on the charge of the ionized dopants in the depletion region [25].

These fluctuations were anticipated long ago [26, 31] and have always been important for SRAM bitcells and analog circuits, due to their sensitivity to mismatch [26, 31]. With technology scaling, the number of dopants in the depletion region has been decreasing steadily, as shown in Fig. 2.7. The decrease has been roughly proportional to L , so that we are now into the regime in which the smallest FETs have less than few hundred dopants determining the threshold voltage [32]. Following Poisson statistics, fluctuations in the dopant number have a standard deviation equal to the square root of the number of dopants, which causes a large increase in V_{th} variation as shown in Fig. 2.7.

Fig. 2.7 Impact of RDF on σV_{th} and number of dopant atoms in the depletion layer of a MOSFET for different technology nodes [32, 33]



The pioneering work of [26, 27, 31] showed that the variation in V_{th} due to random dopant fluctuations (RDF) follows a Gaussian distribution, and its standard deviation can be modeled as:

$$\sigma V_{th} = \left(\sqrt[4]{2q^3 \epsilon_{Si} N_a \phi_B} \right) \times \frac{T_{ox}}{\epsilon_{ox}} \times \frac{1}{\sqrt{3} W L} \quad (2.1)$$

where q is the electron charge, ϵ_{Si} and ϵ_{ox} are the permittivity of the silicon and gate oxide, respectively, N_a is the channel dopant concentration, ϕ_B is the difference between Fermi level and intrinsic level, T_{ox} is the gate oxide thickness, and W and L are the channel width and channel length for the transistor, respectively.

Equation (2.1) shows that σV_{th} is inversely proportional to the square root of the active device area. Hence, the transistors can be sized up to mitigate variations, which is one of the main techniques used in analog design to reduce mismatch between transistors [34]. Moreover, V_{th} variation is largest in SRAM devices, which typically use the smallest sizes in a technology. In addition, Eq. (2.1) shows that variation increases with technology scaling. Figure 2.5 shows the large increase in σV_{th} with technology scaling for SRAM devices. Relative variation can reach about 50% of V_{th} in advanced technologies which has strong impact on SRAM operation.

2.3.2 Line Edge Roughness

Gate patterning introduces a non-ideal gate edge; this imperfection is referred to as LER, as shown in Fig. 2.8. As device scaling continues into sub-50 nm regime, LER is expected to become a significant source of variation due to its direct impact on σV_{th} [24, 35, 36].

Figure 2.9 shows that nanometer technologies use light sources with wavelengths which are much larger than the minimum feature size [17] which increases gate variation due to LER. In addition, the patterning of features smaller than the wave-

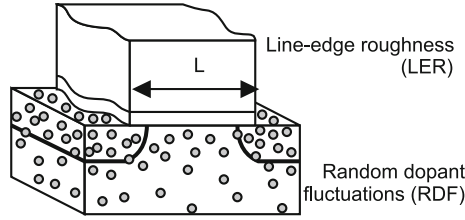


Fig. 2.8 Primary sources of variation in nanometer technologies: RDF and LER [33]

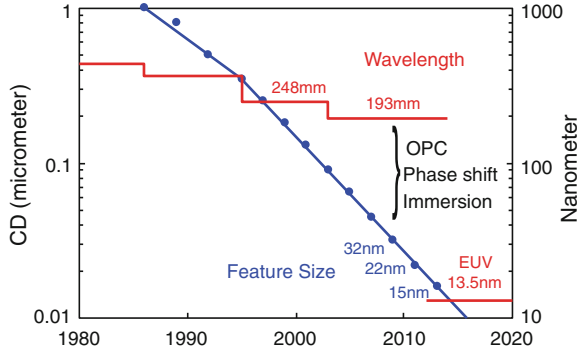


Fig. 2.9 Lithography wavelength scaling for different technology nodes critical dimension (CD). Recent technology nodes used 193 nm optical lithography with enhancements such as optical proximity correction (OPC), aperture improvement, and immersion technology to extend the lithography capabilities down to 15 nm [36]

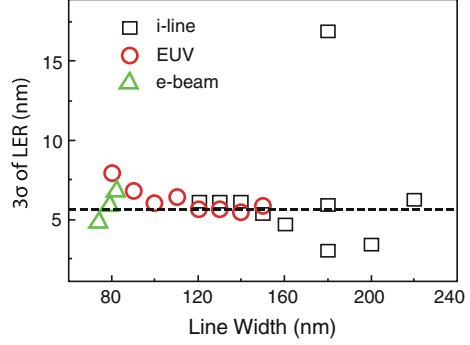
length of light used in lithography results in distortions due to the diffraction of light, referred to as optical proximity effects (OPE) [12, 36, 37]. OPEs cause large variations in defining the minimum feature sizes (e.g., critical dimension CD), and increases LER variations [15, 36, 38].

Controlling LER variations is extremely difficult since the variations do not scale with technology; in other words, the improvements in the lithography process does not reduce LER. Figure 2.10 shows LER for different lithography technologies illustrating that LER variation is almost constant for different technology nodes, which means that for shorter channel lengths, the impact of LER is larger.

The variation in transistor channel length due to LER has a direct impact on several electrical properties of a transistor; however, the most affected parameters are the transistor drive current ($I_D \propto 1/L$) and V_{th} [16, 25]. The variation in V_{th} arises due to the exponential dependence of V_{th} on channel length L for short channel devices, mainly due to short-channel effects (SCE) and the drain-induced barrier lowering (DIBL) [16, 25]. Both effects cause V_{th} to change strongly dependent on the channel length L as shown in Fig. 2.11. V_{th} shift due to SCE and DIBL can be modeled as [16, 25]:

$$V_{th} \approx V_{th0} - (\zeta + \eta V_{DS})e^{-L/\lambda} \quad (2.2)$$

Fig. 2.10 Amplitude of line edge roughness (LER) versus line width for different lithography technologies [33]



where ζ is the SCE coefficient, η is the DIBL effect coefficient, and V_{th0} is the long channel threshold voltage. Therefore, a slight variation in channel length will introduce large variation in V_{th} , as shown in Fig. 2.12.

This type of variation strongly depends on the applied drain-to source-voltage V_{DS} , and the body bias V_{BS} , as shown in Fig. 2.13, because DIBL strongly depends on both V_{DS} and V_{BS} voltages [16, 25]. The roll-off increases as V_{DS} increases. Moreover, as shown in the figure, V_{th} roll-off decreases when forward biasing the body (i.e., V_{BS} positive for NMOS), and vice versa for reverse body biasing (RBB). Therefore, the impact of L variation on V_{th} reduces when applying forward body bias (FBB) [16, 25].

From a circuit modeling approach, the total variation in V_{th} due to RDF, LER, and other static sources of variation, can be formulated as:

$$\sigma_{V_{th}}^2 \approx \sigma_{V_{th,RDF}}^2 + \sigma_{V_{th,LER}}^2 + \sigma_{V_{th,other}}^2 \quad (2.3)$$

2.3.3 Random Telegraph Noise

As transistors continue to become smaller, the impact of single charge perturbation becomes more significant, leading to increased RDF as discussed in earlier sections. In addition to RDF other types of variations arise such as random telegraph noise (RTN), also known as random telegraph signal, RTS [40, 41]. RTN is a random fluctuation in device drain current due to the trapping and detrapping of channel carriers in the dielectric traps at the oxide interface, as shown in Fig. 2.14, which causes variation in V_{th} . The fluctuation in drain current is caused by the change in the number of carriers as well as the changes in surface mobility due to scattering by the trapped charges in the gate dielectric [16]. Both RTN and RDF arise due to discreteness in charges; however RTN significantly differs from RDF in that it is time-dependent, and much fewer charges are involved [40]. Dealing with RTN

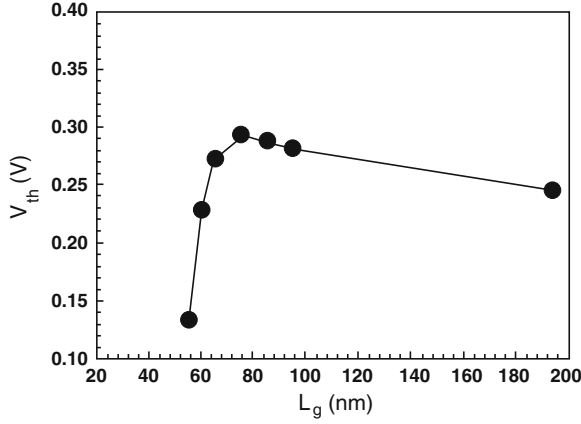
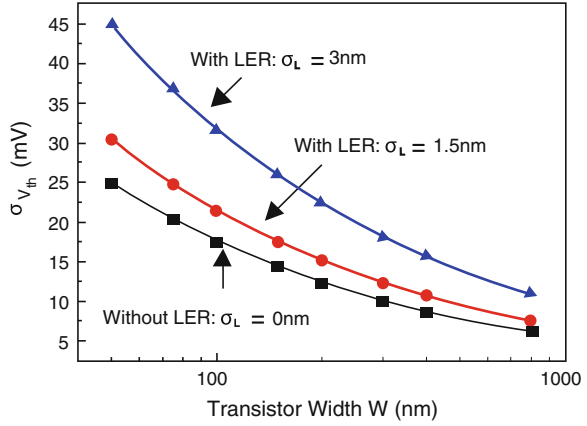


Fig. 2.11 V_{th} versus channel length L in a 90 nm technology [39]. Measured data show the impact of strong short channel effects (SCE) on V_{th} for short channel devices

Fig. 2.12 Combined effect of RDF and LER on V_{th} variation. LER increases the variation of V_{th} beyond the values expected from RDF only ($\sigma_L = 0$ nm). Results are based on 65 nm predictive models [33]



noise has always been important for analog and radio frequency circuits [34]. With technology scaling, RTN increases due to reduction in the number of channel carriers caused. RTN may become a serious issue for SRAM in the near future [40, 41].

The impact of RTN on V_{th} variations can be estimated as follows:

$$\Delta V_{th,RTN} = \frac{q}{W_{eff}L_{eff}C_{ox}} \quad (2.4)$$

where q is the elementary charger, L_{eff} and W_{eff} are the effective channel length and width, respectively, and C_{ox} is the gate capacitance per unit area. Equation (2.4) shows that $\Delta V_{th,RTN}$ is inversely proportional to the device area, and can therefore become a serious concern for highly scaled technologies. As shown earlier, $\Delta V_{th,RDF}$ variation due to RDF is inversely proportional to the square root of device area, while

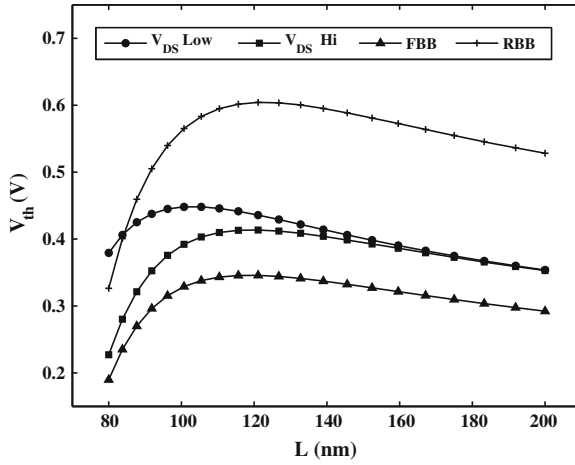


Fig. 2.13 Simulated V_{th} versus channel length L showing V_{th} roll-off under low and high V_{DS} and forward (FBB) and reverse body bias (RBB)

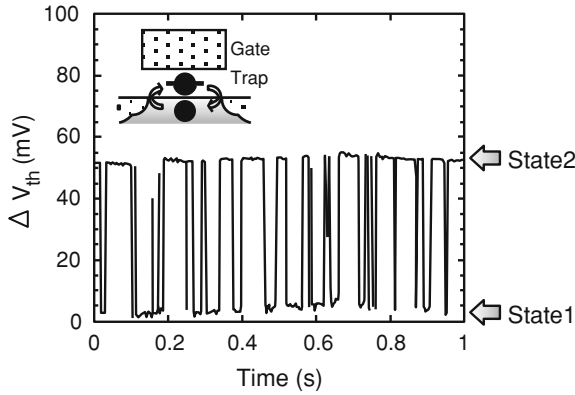


Fig. 2.14 Random telegraph noise (RTN) V_{th} variation is caused by trapping and detrapping of charges in the channel [41]

$\Delta V_{th,RTN}$ is inversely proportional to device area. Therefore, with technology scaling, variation due to RTN is expected to exceed the RDF component [29, 41, 42].

Detailed measurements in 22 nm generation have shown that $\Delta V_{th,RTN}$ exceeds 70 mV for smaller devices [41]. One of the critical concerns with RTN is that V_{th} variation due to RTN has a non-Gaussian distribution with a long tail, as shown in Fig. 2.15. RTN V_{th} variations are projected to exceed RDF V_{th} beyond the 3σ point; hence, RTN may exceed RDF in design impact [41].

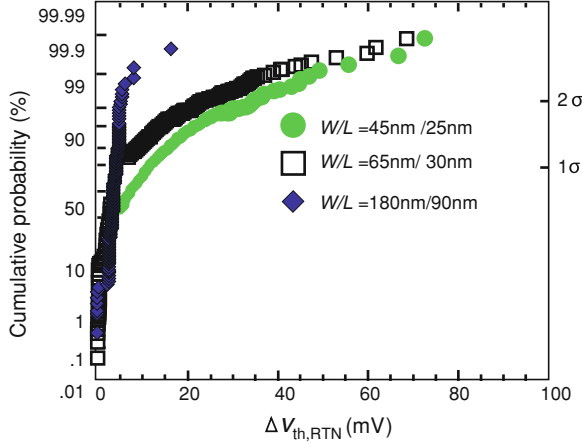


Fig. 2.15 Distribution of V_{th} fluctuation due to random telegraph noise (RTN) in a 22 nm technology, which shows a long-tailed distribution [41]

2.3.4 Time-Dependent Degradation and Aging

In addition to static (time-independent) sources of variations and intrinsic device noise such as RTN, highly scaled devices are degraded over time due to long-term stress and device aging. One of the most critical sources of device aging is negative-bias temperature instability (NBTI), which affects PMOS devices. NBTI causes the magnitude of V_{th} to shift due to the generation of interface traps at the oxide interface when a negative voltage is applied to the PMOS gate for a long period of time. NBTI threshold voltage shift $\Delta V_{th,NBTI}$ due to static stress after time t can be modeled as [43]:

$$\Delta V_{th,NBTI} = A((1 + \delta)t_{ox} + \sqrt{C(t - t_0)})^{2n} \quad (2.5)$$

$$A = \left(\frac{qt_{ox}}{\epsilon_{ox}}\right)^{\frac{1}{2n}} \sqrt{K^2 C_{ox} (V_{gs} - V_{th}) \exp\left(\frac{V_{gs} - V_{th}}{t_{ox} E_0}\right)} \quad (2.6)$$

$$C = \frac{\exp(-E_a/kT)}{T_0} \quad (2.7)$$

where n is the time exponent (1/6 or 1/4 depending on the NBTI model), q is the electron charge, T is the temperature, k is the Boltzmann constant, and C_{ox} is the oxide capacitance per unit area. A is proportional to the hole density and depends exponentially on temperature, and T_0 , E_a and E_0 , K are constants that are extracted by fitting the measured data. Equation (2.7) shows that $\Delta V_{th,NBTI}$ shift strongly depends on temperature (via C) and oxide thickness t_{ox} (via A) [43]. Due to the reduction of t_{ox} with technology scaling, the impact of NBTI on V_{th} increases.

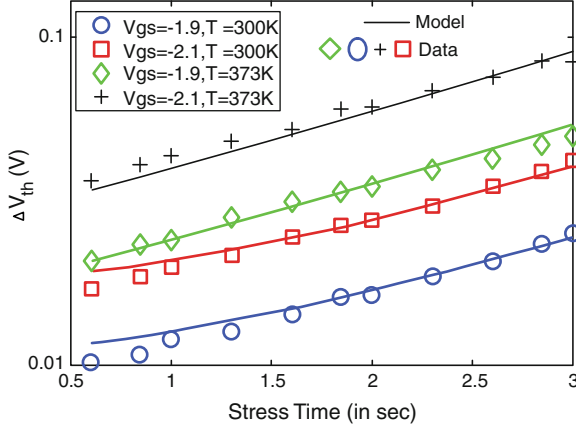


Fig. 2.16 Measured V_{th} degradation under static NBTI for different temperatures and V_{gs} for 90 nm technology [43]

V_{th} shift due to NBTI can be simplified as:

$$\Delta V_{th,NBTI} \approx At^n \quad (2.8)$$

where n is the time exponent which has been shown to vary widely (0.14–0.5) [44, 45].

Figure 2.16 shows V_{th} degradation under static NBTI for 90 nm technology at different temperature and voltage conditions. NBTI shift recovers slightly after the stress condition is removed. Models for V_{th} shift that account for recovery and dynamic stress have been developed. It is important to note that recovery time is in the mS to S range, which complicates the process of accurately measuring the NBTI stress effects [46].

For newer technologies using high-K dielectrics, NMOS devices suffer from a similar reliability concern due to positive bias temperature instability (PBTI) [47]. In addition, there are other sources of time-dependent device degradation due to aging such as the hot carrier effect (HCI) [16, 25] and time-dependent dielectric breakdown (TDDB).

2.3.5 Other Sources

While random dopant fluctuation and LER are currently the dominant sources of device variations, there are several other sources which may become significant in the future technologies. Below we list other sources of device variations:

- **Oxide Charges Variation:** Interface charges can also cause V_{th} variation although their effect is not significant in nitrided gate oxides [24]. The recent adoption of

high-K gates to reduce gate-tunneling leakage current may worsen oxide charge variations [24]. In addition, oxide charge variations can introduce mobility fluctuations, as they increase scattering in a transistor channel.

- **Mobility Fluctuation:** Variations in a transistor's drive current can also be caused by mobility fluctuation. Mobility fluctuation can arise from several complex physical mechanisms such as fluctuations in effective fields, fixed oxide charges, doping, inversion layer, and surface roughness [24]. Throughout their shared dependence on many physical variation mechanisms, mobility variation shows a certain level of correlation with V_{th} variations. Device measurements show this correlation to be small [48]. Therefore, mobility variations and V_{th} variations are typically assumed to be independent in circuit modeling [48].
- **Gate Oxide Thickness Variation:** Any variation in oxide thickness affects many electrical parameters of the device, especially V_{th} . However, oxide thickness is one of the most well-controlled parameters in MOSFET processing. Therefore, it does not affect V_{th} variation significantly.
- **Channel Width Variation:** Due to lithography limitations, transistor channel width also varies due to lithography (similar to LER). Width variations can cause V_{th} variations in devices that suffer from narrow-width effects (NWE) [16]. However, since W is typically 2–4 times larger than L , the impact of W variation on V_{th} is considered to be smaller than the impact due to L variation [16].

2.3.6 Highly Scaled Devices: FinFET

At the time of writing this book, FinFET or trigate transistors started showing up in production [49, 50]. FinFET is considered a fundamental change in CMOS technology because the device moved from being planar to becoming a 3D structure as shown in Fig. 2.17. FinFET technology reduces SCE and significantly reduces leakage current, which allows lower operating voltage [49, 50]. Due to the improved electrostatic integrity in FinFET, the channel control via the gate can be maintained with lower channel doping. As the fin thickness is reduced, the FinFET becomes fully depleted, the channel can be undoped which provides large reduction in V_{th} fluctuation due to RDF. Figure 2.18 shows a comparison between V_{th} variation for 45 nm bulk and FinFET. With a 6X reduced channel doping, FinFET can reduce V_{th} variations by more than 60%.

While FinFET structure helps reduce the RDF contribution to variation, it adds new sources of variation such as fin thickness variations, as shown in Fig. 2.19. Moreover, the FinFET width is determined using fin height, which is constant for a given technology, so the transistor width is quantized and cannot be adjusted in fine steps as in bulk technologies. Width quantization is a major concern for SRAM bitcell design since the current ratios of pass-gate (PG), pull-down and pull-up can only take integer values as shown in Fig. 2.17, preventing optimization of read and write stability through sizing as in conventional bulk technologies. Therefore, read and write assist techniques are necessary for FinFET SRAM [49, 50].

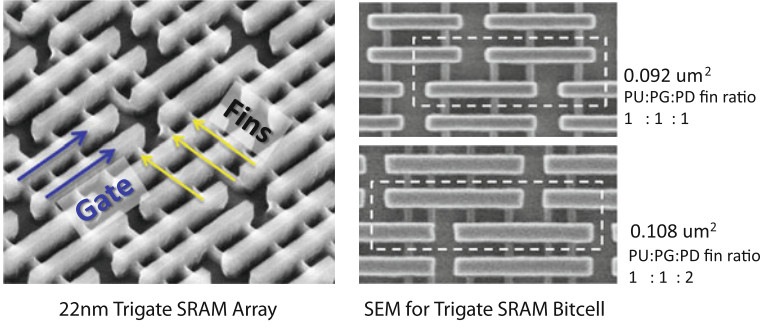


Fig. 2.17 22 nm FinFET SRAM array and the layout of two different FinFET bitcells. Width quantization limits the ratios of pull-up, pass-gate, and pull-down devices to integer values [49, 50]

Fig. 2.18 Comparison between V_{th} variation due to RDF for conventional bulk technology (45 nm) and FinFET (trigate) using a lightly doped channel. V_{th} variation reduces significantly for FinFET [51]

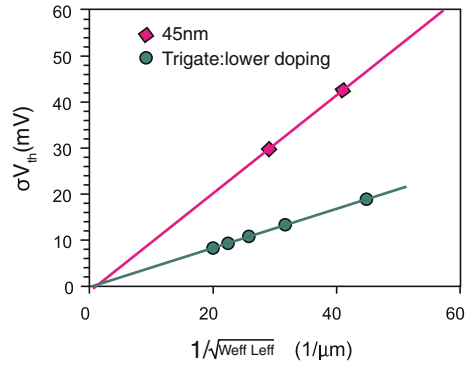
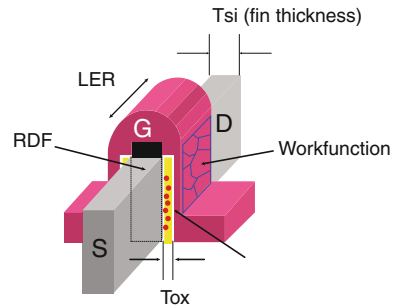


Fig. 2.19 Sources of V_{th} variation in highly scaled FinFET including RDF, LER, fin thickness (T_{si}), oxide thickness (T_{ox}), and workfunction variations [52]



2.4 Interconnect Variability

Similar to the sources of variations that alter device characteristics, several factors affect the interconnects. The main sources of variations in interconnects include [14]:

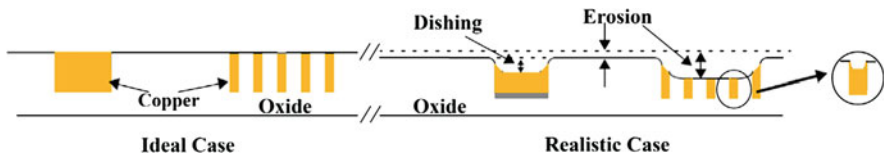


Fig. 2.20 A schematic cross-section of interconnect showing dishing and erosion impact on metal height

1. **Line Width and Line Space:** Deviations in the width of patterned lines arise primarily due to photolithography and etch dependencies. At the smallest dimensions, which typically occur at lower metal levels, proximity and lithographic effects are likely most important. However, at higher metal levels, aspect ratio-dependent etching, which depends on line width and local layout, are generally more significant. Variations in line width directly impact line resistance as well as line capacitance [14, 15].
2. **Metal and Dielectric Thicknesses:** In a conventional metal interconnect, the thickness of metal films is usually well controlled, but can vary from wafer-to-wafer and across the wafer. However, in advanced damascene copper interconnect processes, this is not the case. Unlike older aluminum interconnect processes where the metal is patterned and the oxide is polished, the oxide is patterned, and the metal is polished in a damascene process for copper interconnects. Chemical mechanical polishing (CMP) is then used to flatten the topography on the wafer. Because copper and adjacent dielectric are removed from the wafer at different rates during CMP (depending on the density of the surrounding pattern), this creates surface anomalies such as dishing and erosion. Dishing occurs when the copper recedes below the level of adjacent dielectric and erosion is a localized thinning of the dielectric, which normally happens when CMP is applied to an array of dense lines, as shown in Fig. 2.20. The oxide between wires in a dense array tends to be over-polished compared to the nearby areas of wider insulators. Dishing and oxide erosion are layout dependent; they are problematic in wide lines and dense arrays, respectively. In damascene processes with copper interconnects, dishing and erosion can significantly impact the final thickness of patterned lines, with line thickness losses of 10–20 % leading to higher resistance and capacitance variations [14, 15].
3. **Contact and Via Size:** Contact and via sizes can be affected by variations in the etching process, as well as layer thickness variations. Depending on the via or contact location, the etch depth may be substantially different, resulting in different sizes of the lateral opening. Such size differences can directly change the resistance of the via or the contact [14].

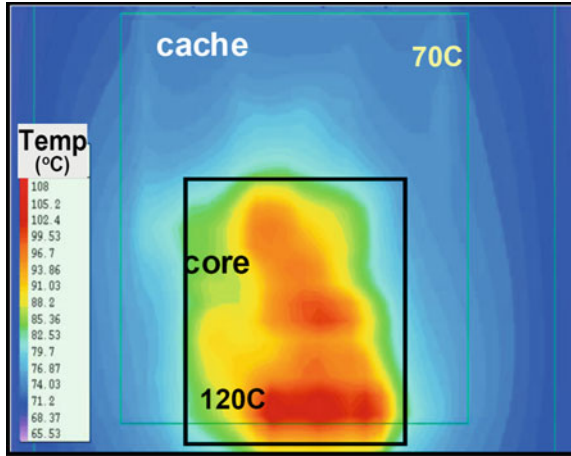


Fig. 2.21 Thermal image showing within die temperature variation for a microprocessor [18]. Hot spots with temperatures as high as 120°C are shown

2.5 Environmental Variability

In addition to static process variations, environmental factors, which are typically dynamic, also cause variation in the circuit operation. These include variations in power supply and temperature of the chip or across the chip [13, 17, 19].

Variation in switching activity across the die results in uneven power dissipation across the die and uneven supply voltage noise (droops or bumps). A reduced power supply lowers drive strengths, degrades speed for digital logic, and increases SRAM failures [18, 53].

WID temperature fluctuations have always been a major performance and packaging challenge, especially for high-performance processors. Temperature fluctuations are problematic because both device and interconnect have temperature dependencies which cause performance to degrade at higher temperatures. Moreover, temperature variation across communicating blocks on the same die can cause performance mismatches, which may lead to functional failures [18]. As shown in Fig. 2.21, the temperature difference between the core and the cache of a microprocessor can be as high as 50°C.

Leakage currents, especially subthreshold leakage, strongly depend on temperature; leakage power increases at higher temperatures [14, 17, 18]. In the meantime, higher leakage power causes die temperature to rise. This type of positive feedback may cause thermal runaway where leakage currents, and temperature continue to increase until failure [15].

Both supply and temperature variations depend on the work load of the processor and are thus, time-dependent. However, identifying worst-case conditions for temperature and supply is very difficult [15]. Therefore, designers often focus on

minimizing temperature and supply variations as much as possible; for example, ensuring that the voltage drop on the power grid is always less than 10% of the nominal supply voltage, and by adding large decoupling capacitors [15, 17].

2.6 SRAM Failure Mechanisms

Due to their small size and high density, several factors can cause SRAM bitcells to fail. Figure 2.22 shows various SRAM failures such as stability fails, radiation induced soft errors, and hard fails.

2.6.1 Bitcell Stability Failures

The effect of technology scaling on SRAM failure probability is shown in Fig 2.23 for advanced technology nodes spanning 130 nm down to 45 nm. Traditional hard fails due to defect density decrease due to the reduction of bitcell size and improvement in defect density. However, as the bitcell size is reduced by about 50% in each technology node, process variations increase significantly and become the dominant cause of bitcell failure [3, 6, 7]. This increase in SRAM failures has a strong impact on the overall product yield due to the high memory densities on chip. Moreover, lower V_{DD} operation becomes limited by the SRAM minimum supply voltage V_{min} due to the sensitivity of stability failures to supply voltage.

There are four main parametric failure mechanisms (also known as SRAM stability failures) [5, 9, 54–56]:

1. read access failure;
2. read stability or read disturb failure;
3. write failure;
4. hold or retention fail.

These failures are parametric in nature since they affect the memory operation under specific conditions. For example, these failures mostly appear as low V_{DD} , while they can be recovered at higher supply voltages. Therefore, these failure mechanisms become the limiting factor for SRAM supply voltage scaling [44, 57, 58].

2.6.1.1 Read Access Failure

During read operation, the wordline (WL) is activated for a small period of time determined by the cell read current, bitline loading (capacitance) as shown in Fig. 2.24. The content of a cell is read by sensing the voltage differential between the bitlines using a sense amplifier. For successful read operation, the bitlines precharged to V_{DD} should discharge to a voltage differential value which can trigger the sense amplifier correctly. Read failure occurs if bitcell read current (I_{read}) decreases below a certain limit, which often results from an increase in V_{th} for the PG or pull-down (PD)

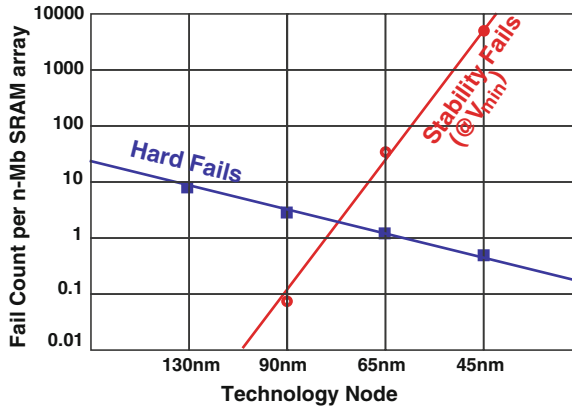


Fig. 2.23 SRAM hard fails and stability fails versus technology scaling [7]

2.6.1.2 Read Stability Failure

SRAM cells are designed to ensure that the contents of the cell are not altered during read access, and the cell can quickly change its state during write operation. These conflicting requirements for read and write operations are satisfied by sizing the bitcell transistors to provide stable read and write operations [5, 9, 54].

In read operation, an SRAM bitcell is most prone to failure. After the WL is enabled, voltage at the internal storage node storing a zero (Q) slightly rises due to the voltage divider between the PG transistor (PG1) and the pull-down (PD1), as shown in Fig. 2.24. If the voltage at Q rises close to the threshold voltage of the adjacent pull-down, PD2, the cell may flip its state. Therefore, stable read operation requires that PD1 should be stronger than PG1. Read stability failure is exacerbated by process variations, which affect all the transistors in the bitcell [5, 9, 54]. To quantify the bitcell's robustness against this type of failure, static noise margin (SNM) is one of the most commonly used metrics [61]. A read stability failure can occur if the bitcell cannot hold the stored data, in which case SNM is zero [5, 54, 61].

Read stability failure can occur any time the WL is enabled even if the bitcell is not accessed for either read or write operations. For example, in half-selected bitcells, the WL is enabled while the bitlines column is not selected (the bitcells are not actively accessed for read or write). These bitcells experience a dummy read operation because the bitlines are initially precharged to V_{DD} , and the bitlines are discharged after the WL is enabled, hence, the bitcells become prone for read stability failure. Dealing with read stability failures is one of the biggest challenges for SRAM design and has been extensively studied [5, 9, 54]. Circuit techniques to deal with read stability failures will be discussed in Chap. 3.

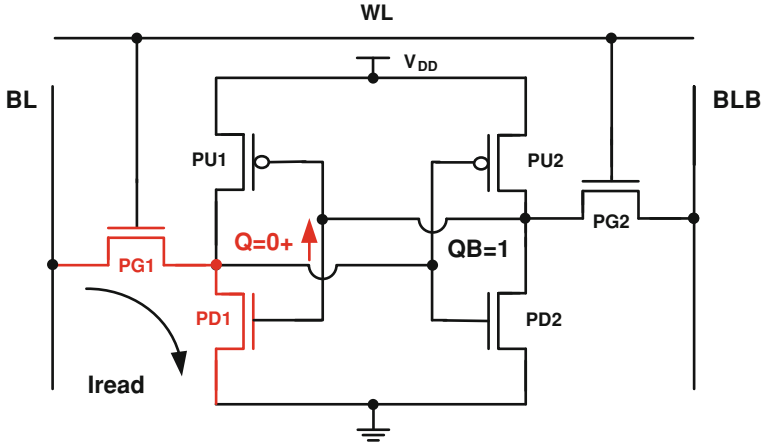


Fig. 2.24 Bitcell in read operation

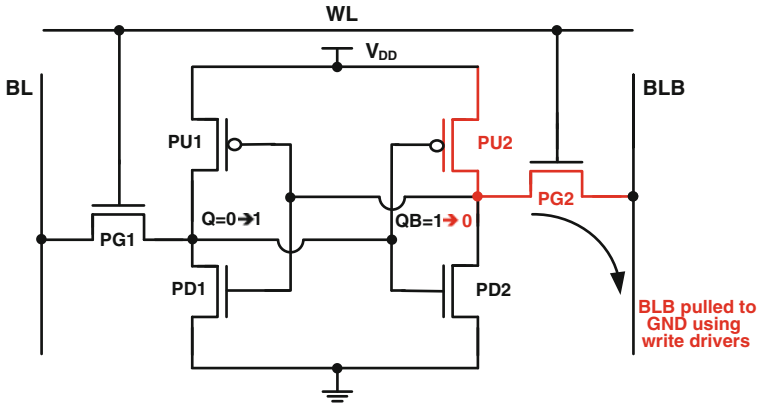


Fig. 2.25 Bitcell in write operation

2.6.1.3 Write Stability Failure

The ability of the bitcell to be written correctly is referred to as write stability or write margin. In write operation, BLB is pulled to zero by the write driver, while WL is enabled, as shown in Fig. 2.25. Therefore, the NMOS PG2 is turned ON, which results in a voltage drop in the storage node QB holding data 1 until it falls below $V_{DD} - V_{th}$ for the PMOS PU1, where the positive feedback action begins. For stable write operation, PG2 should be stronger than PU2. Due to WID variations, the pass gate cannot overcome the pull-up transistor, resulting in a write failure [54, 62]. Write failure can also happen if the WL pulse is not long enough for the bitcell to flip the internal nodes (dynamic failure).

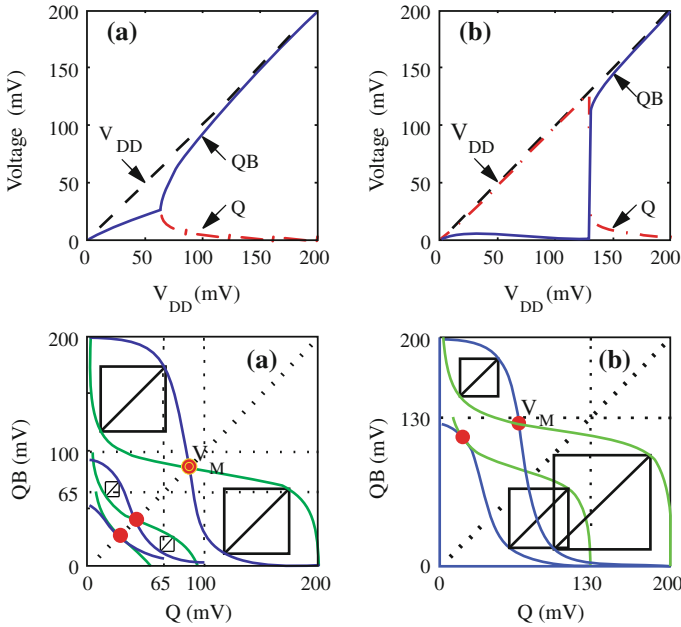


Fig. 2.26 Data retention failure mechanism. *Upper figures* show the bitcell internal node voltages Q and QB for **a** balanced and **b** imbalanced cell as V_{DD} is reduced. *Lower figures* show the voltage transfer characteristics (VTC) of **a** balanced and **b** imbalanced cell with varying V_{DD} . V_M is the trip point of the VTCs. The imbalanced bitcell has higher data retention voltage (DRV) than the balanced case because of asymmetry in the VTCs caused by variations [58]

2.6.1.4 Data Retention Failure

Reducing supply voltage (V_{DD}) is one of the most effective techniques to reduce both static and dynamic power consumption for digital circuits [63]. In SRAM, the data retention voltage (DRV) defines the minimum V_{DD} under which the data in a memory is still preserved. When V_{DD} is reduced to DRV, all six transistors in the SRAM cell operate in subthreshold region, and therefore are strongly sensitive to variations [57, 58].

DRV depends strongly on WID variations in the bitcell inverters, which may cause the bitcell to be imbalanced. This imbalance can be examined using SNM in standby (WL is disabled) as shown in Fig. 2.26. If the bitcell is asymmetric due to WID variations, the bitcell tends to have a higher DRV than in the symmetric case. This can be explained using SNM, where DRV voltage can be defined as the voltage when hold SNM is equal to zero. In the symmetric case, both SNM high (upper left square) and SNM low (lower right square) decrease symmetrically to zero. However, in the case of asymmetric bitcell shown in Fig. 2.26, SNM low is always larger than SNM high, and the bitcell DRV is limited by the SNM high case. Therefore, variations increase the bitcell DRV because they increase the asymmetry [57, 58].

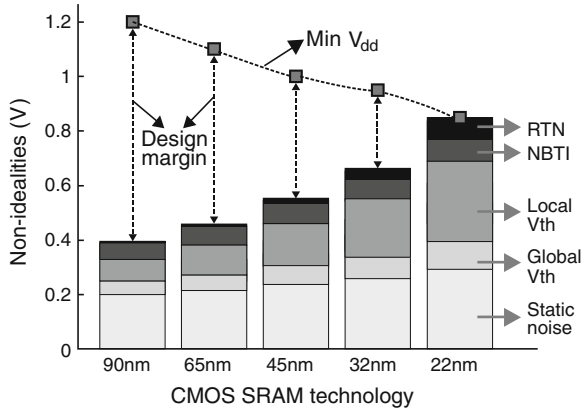


Fig. 2.27 Reduction of SRAM design margin with technology scaling due to increase in variations [64]

2.6.2 Impact of Variations on SRAM V_{\min}

Variations affect SRAM operations and set the limit of the SRAM minimum supply voltage V_{\min} . Due to the increase in variations with technology scaling, the operation margin of SRAM decreases, as shown in Fig. 2.27, making the SRAM V_{\min} the limiting factor of voltage scaling in newer technologies. Measured distribution of SRAM V_{\min} is shown in Fig. 2.28. The distribution shows the large spread of V_{\min} due to the impact of RDF and LER on the bitcell stability. Figure 2.29 shows the impact of WID variations on SNM, which translates into larger V_{\min} spread.

As discussed earlier, NBTI and PBTI cause V_{th} of PMOS and NMOS devices to increase, which decreases the bitcell stability [66–68]. Figure 2.30 shows that as SNM decreases due to NBTI, the probability of read disturb increases. The increase in read disturb failures causes an increase in the V_{\min} since higher voltage is required to recover the degradation in SNM. Figure 2.31 shows the shift in SRAM minimum supply voltage V_{\min} due to burn-in stress. As the burn-time increases, V_{\min} shifts towards higher voltage, which can be attributed to the increase in PMOS V_{th} due to NBTI and the associated reduction in SNM. For some parts, stress causes V_{\min} to decrease, as shown in the negative values of V_{\min} shift, which may be explained by improvements in bitcell write margin since PMOS pull-up becomes weaker. However, even for read-limited memories, V_{\min} shift may be negative; and there is little correlation between initial V_{\min} and its final value after stress, which is due to the combined impact of static variations (RDF, LER) and NBTI degradation [44].

Depending on the bitcell size and stability optimization, NBTI's impact on V_{\min} can vary. Figure 2.33 shows that for read-limited bitcells, NBTI causes the V_{\min} distribution to move towards higher voltage. However, for write-limited bitcells, the V_{\min} distribution to lower voltage, since the strength of the pull-up device decreases due to NBTI, which improves write margin [44, 67, 69].

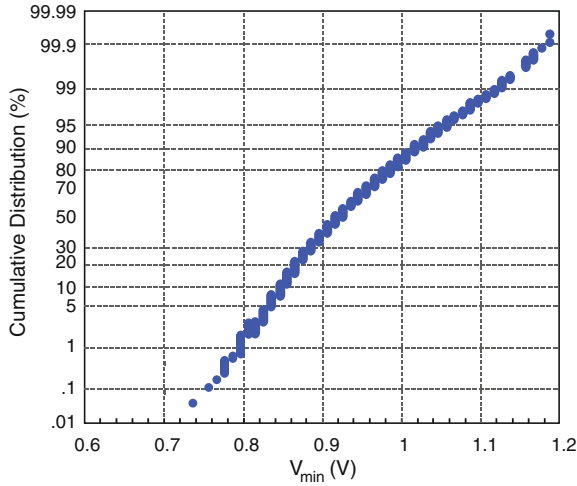
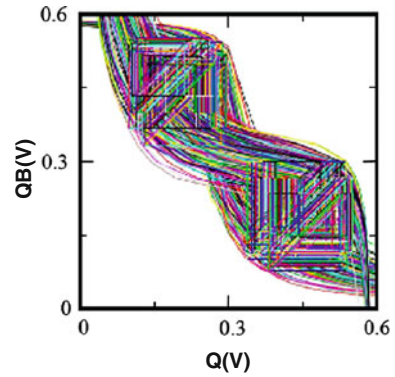


Fig. 2.28 Measured V_{\min} distribution for a 2Mbit SRAM in 65 nm technology. V_{\min} varies widely due to random variations, causing read and write stability failures [65]

Fig. 2.29 Measured SNM butterfly curves for 512 bitcells in a 65 nm technology node showing the strong impact of WID variations on SNM [6]. A large spread in butterfly curves causes SNM to be unsymmetrical and increases the probability of bitcell failure



Another important aspect of NBTI degradation is time dependence and its impact on V_{\min} . As shown in Fig. 2.32, increasing burn-in increases NBTI shift, which increases V_{\min} . The shift in PMOS pull-up V_{th} due to NBTI causes the SRAM V_{\min} to increase linearly [47]. V_{\min} increases by 110 mV after the first 8 h of stress, and increases by an additional 30 mV after 160 h. The initial large shift and eventual saturation of V_{\min} reflects the time dependence of $\Delta V_{th,NBTI}$ which follows a fractional power law $\Delta V_{th,NBTI} \propto t^n$, where $n < 0.5$ as explained earlier [44, 45, 67]. Simulation studies for high-K dielectrics show that SRAMs are more susceptible to V_{\min} stability problems due to the combined effects of PMOS NBTI and NMOS PBTI. Guard-banding or screening for time-dependent shift in SRAMs is a challenge.

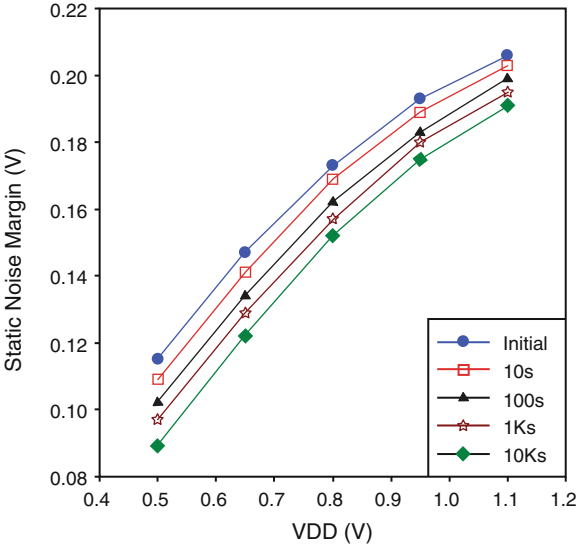


Fig. 2.30 Measured SNM versus V_{DD} for different stress times in a 65 nm node. SNM monotonically decreases with stress time due to NBTI degradation [67]

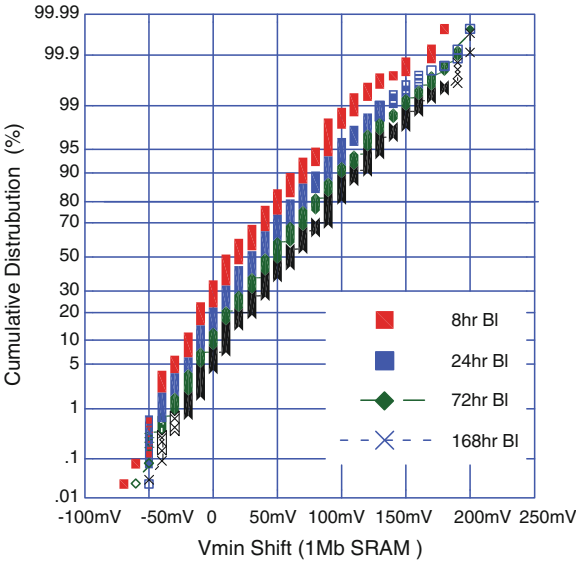


Fig. 2.31 Measured V_{min} shift distribution of a 1 Mb SRAM for different stress/burn-in durations in a 65 nm technology [70]

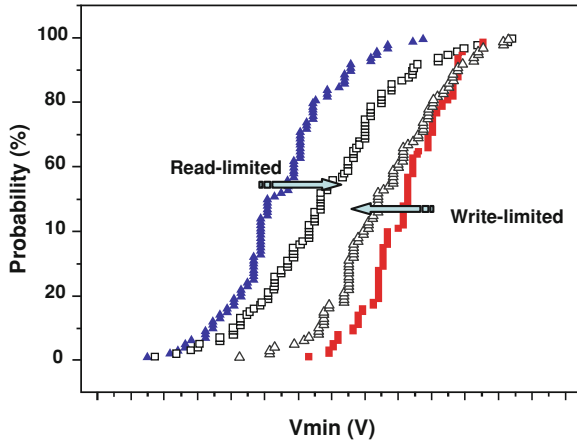


Fig. 2.32 Minimum SRAM supply voltage (V_{min}) distribution for read-limited and write-limited bitcells before (solid markers) and after NBTI stress (hollow markers) [44]

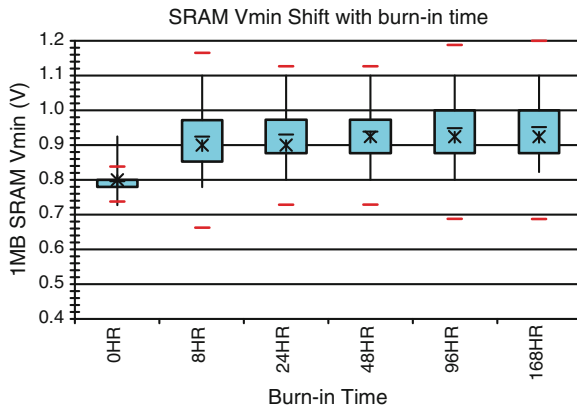


Fig. 2.33 Measured V_{min} shift versus burn-in time in a 65 nm technology [70]

The impact of RTN V_{th} variation on SRAM operation and V_{min} has been researched extensively in the last few years [40, 41, 71–76]. As shown in Fig. 2.34, in smaller bitcells, RTN causes a greater increase in V_{min} shift due to larger $\Delta V_{th,RTN}$, as expected from Eq. (2.4). RTN can increase V_{min} by 50–100 mV for an error-free lifetime condition [74, 75] which is a significant portion of the SRAM V_{min} budget. However, other research predicts that RTN V_{min} degradation becomes less significant in larger arrays [75]. Due to the time dependence of RTN, it is difficult to screen for RTN failures by ordinary functional tests. Design margining techniques as well as accelerated tests have been proposed to deal with RTN issues [40, 76]. Recent work also suggests that there is a relationship between RTN and NBTI, so, margining techniques need to account for both effects simultaneously [77].

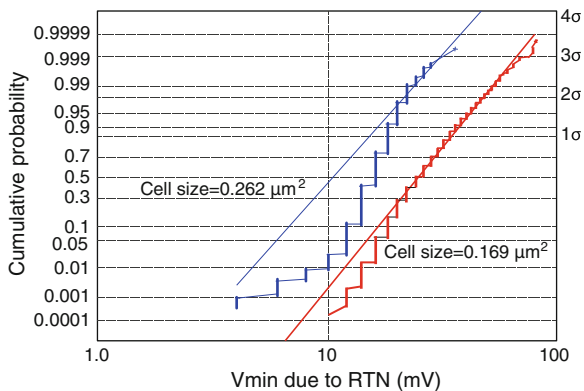


Fig. 2.34 ΔV_{\min} distribution due to RTN for two bitcell sizes, which shows that smaller bitcell has larger V_{\min} shift since RTN is inversely proportional to FET size [74]

2.6.3 Radiation-Induced Soft Errors

SRAMs are susceptible to dynamic disruptions known as single event upsets (SEU) [14]. SEUs arise due to energetic radiation (Alpha particles or cosmic rays) that hits the silicon substrate and generates free electron-hole pairs, which can affect the potential of bitcell storage nodes and flip the stored data. To determine the susceptibility of SRAM to SEUs, the critical charge that can cause a storage node to be disrupted (Q_{crit}) is calculated. However, with technology scaling, SRAM junction capacitance, cell area, and supply voltage are all scaled down. These reductions have opposing effects on Q_{crit} and the collected charges. However, the combined effect causes the rate of SEUs, or SER, to saturate or slightly decrease with technology scaling [78–80], as shown in Fig. 2.35. The reduction in single-bit SER does not necessarily translate into a reduction in the overall system failure rate due to the rapid growth in embedded SRAM density. In fact, SRAM systems failure rates are increasing with scaling and have now become a major reliability concern for many applications [3, 78, 79]. Moreover, process variations lead to large variation in Q_{crit} which also affects SER [81]. Other research, shows that the aging due to NBTI, oxide breakdown, and hot carriers has negligible impact on SER [82].

To mitigate soft errors, several radiation-hardening techniques can be implemented through process technology (e.g., SOI technology), circuit design (e.g., adding feedback capacitors, larger transistors, columns/words interleaving) and architecture (e.g., parity, error correction codes) or a combination of all these techniques [78]. The SER budget for chips or systems is typically set based on target market requirements. For example, for single user, single chip applications as in mobile phones, it is acceptable to have an average failure rate of about one error every two years due to SER. On the other end of applications spectrum, the same

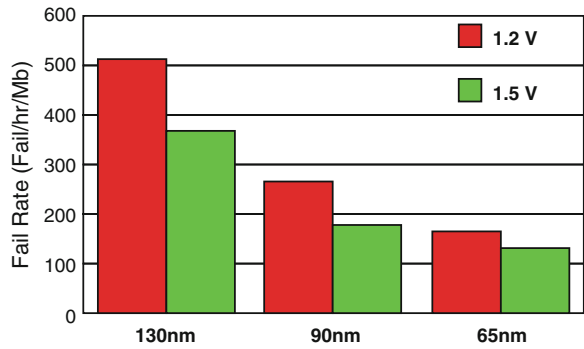


Fig. 2.35 SER failure rate for different technology nodes. SER per Mbit decreases with scaling [3]

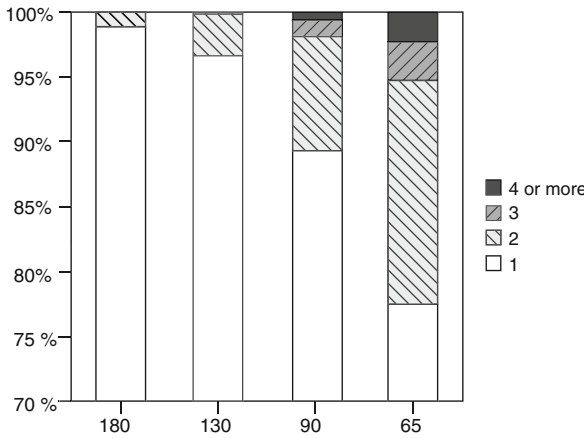


Fig. 2.36 Multicell upsets (MCUs) for different technology nodes, which shows the increase in MCUs with scaling [85]

failure rate is not acceptable for high reliability systems utilizing hundreds of chips as in telecom base stations or servers [78, 79, 83].

In addition to SEUs, multi-cell upsets (MCUs) are becoming a concern for high integrity systems. An MCU consists of simultaneous errors in more than one memory cell and is induced by a single event upset [3, 84, 85]. Risk of MCUs is typically minimized by column interleaving and ECCs. Figure 2.36 shows the increase in MCUs with technology scaling, which results from closer physical proximity of neighboring cells. The increase in MCUs may become a critical reliability concern with future technology scaling [3, 84, 85].

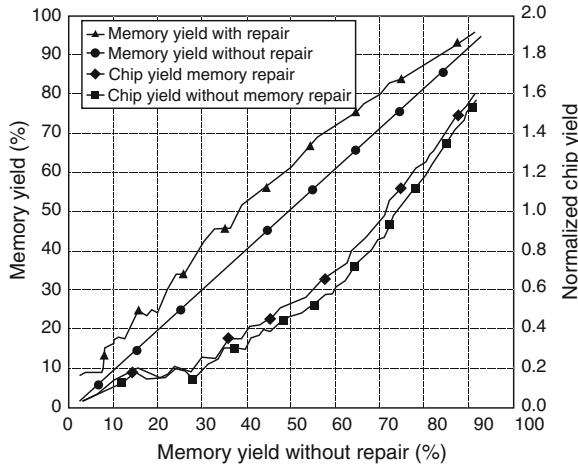


Fig. 2.37 Impact of memory repair on memory and chip yield [88]

2.6.4 Hard (Catastrophic) Fails

Hard fails due to physical defects, also called catastrophic fails, can cause permanent damage for memory and digital circuits. Physical defects include a wide range of possible defects such as voids, shorts, metal bridges, missing contacts or vias, oxide pin holes and many others [14]. Because memories are designed with aggressive design rules, memories tend to be more sensitive to manufacturing defects than other logic circuits in the chip [1]. Figure 2.23 shows that hard fails decrease with process technology due to lower device area, while soft fails due to intrinsic variation increase [3].

SRAM employs redundancy in rows, columns or banks, allowing replacement of defective elements, which improves yield significantly [1, 14, 86]. Historically, this type of repair capability was implemented to address hard fails. However, nowadays, memory redundancy is also used to recover from yield loss due to bitcell stability failures [54, 87].

Figure 2.37 shows how memory repair can improve both memory and chip yield significantly [88, 89]. Memory repair can be used to enhance yield by 5 to 20%, depending on the type of redundancy used, which translates into large improvement in the chip yield. Several yield models have been proposed to analyze the impact of hard defects on memory and chip yield [90, 91]. In addition, yield simulation using inductive fault analysis (IFA) and critical area analysis (CAA) are used to estimate memory yield [90, 91].

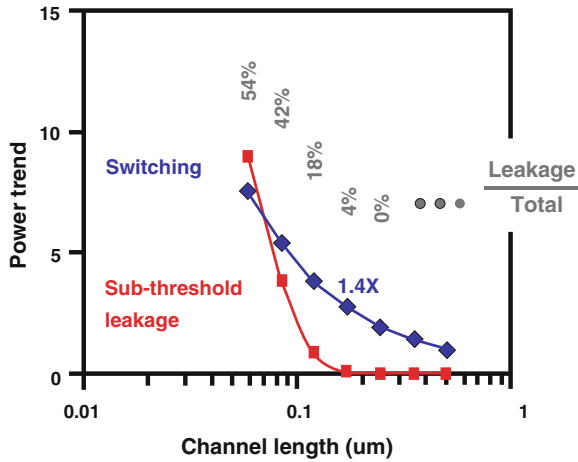


Fig. 2.38 Dynamic and static power versus technology scaling, showing the exponential increase in leakage power in smaller technologies [94]

2.7 Techniques to Deal with Variability for Logic Circuits

In this section, we review state-of-the-art research studying the increase in variability for logic circuits. While the focus of this book is about SRAM, it is instructive to look at the impact of variations on logic circuits to highlight the similarities in the variation-tolerant design approaches.

Performance and power consumption are the most critical metrics for logic circuits. In nanometer devices there are several sources of leakage current, such as subthreshold, gate oxide tunneling, junction band-to-band tunneling (BTBT), and gate-induced drain leakage (GIDL), all of which increase with technology scaling [32, 92, 93]. Therefore, for designs in sub-90nm, leakage is considered a significant part of the total power, and it increases with technology scaling as shown in Fig. 2.38.

The large variability in advanced CMOS technologies increasingly contributes to the total leakage of a chip because leakage depends strongly on process variations [95, 96]. For example, variation in V_{th} introduces a large spread in subthreshold leakage due to the exponential dependence on V_{th} . Similarly, gate-tunneling leakage current is sensitive to oxide variation. The sensitivity of leakage to variations has accentuated the need to account for statistical leakage variations during design [95–97].

For a whole chip, process variations can cause large variations in leakage power. Figure 2.39 shows measured variation of maximum frequency and leakage power for a chip in 130nm technology [20]. The figure illustrates that the magnitude of leakage variation is much larger than frequency variation (5X leakage spread for a 30% variation in chip frequency). The chips running at the highest frequency

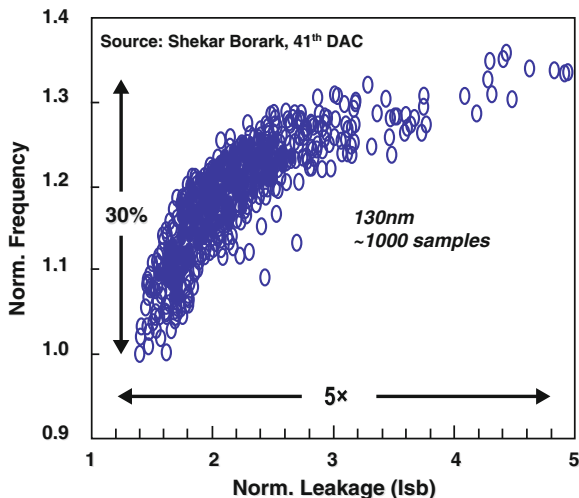


Fig. 2.39 Leakage and frequency variations for a processor in 0.13 μm technology [20]

have a wide distribution of leakage. This excessively large spread in leakage current makes it very difficult to achieve the required speed while meeting power constraints, especially as leakage power increases exponentially with scaling.

Among the chips that meet the required operating frequency, a large fraction dissipate a large amount of leakage power, which makes them unsuitable for usage, degrading the chip yield [18]. The leaky chips have higher frequency because of the inverse correlation between leakage current and circuit delay. For devices with smaller channel length, V_{th} decreases due to short channel effects, and therefore, the subthreshold leakage current increases exponentially. Simultaneously, circuit delay decreases due to the increase in I_{on} , since the overdrive voltage $V_{DD} - V_{th}$ increased. Hence, these chips have a higher operating frequency, but suffer from large leakage power, making them unusable. For the high frequency chips shown in Fig. 2.39, both the mean and standard deviation of leakage current increases considerably, causing yield to decrease substantially [18]. Therefore, there is a crucial need to account for leakage power and its dependence on process variations when analyzing the impact of variability on design techniques [98]. Moreover, variation-tolerant circuit techniques that can reduce variability, and hence reduce leakage power variation, should be designed to improve the yield in advanced CMOS technologies [17, 18, 20].

2.7.1 Circuits

To mitigate variability in logic circuits, considerable research has been done to measure variability and use feedback to mitigate it. These techniques control variability using supply voltage, body bias, or programmable sizing.

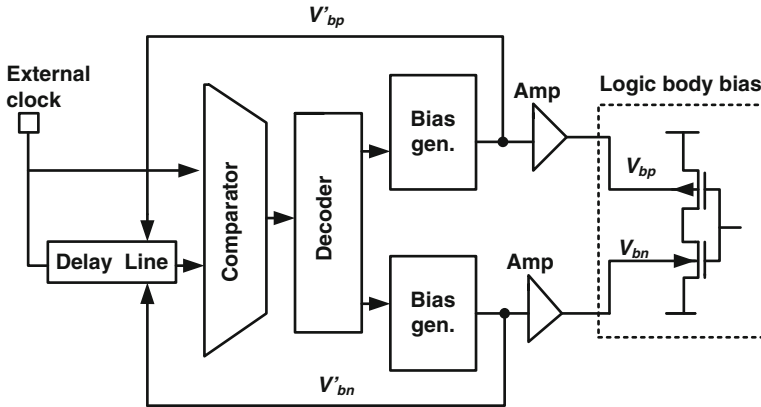


Fig. 2.40 Block diagram of speed-adaptive V_{th} technique [100]

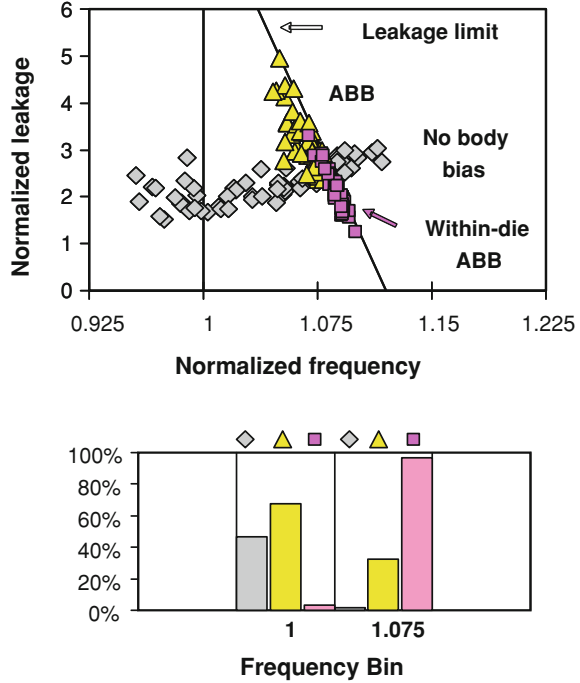
A speed-adaptive body bias technique was utilized in [99, 100] to compensate for variability in microprocessor implemented in $0.2\mu\text{m}$ technology. The speed adaptive- V_{th} is composed of a delay line, a delay comparator, a decoder, and body bias generators, as shown in Fig. 2.40. The comparator measures the delay between an external clock signal and an output signal from the delay line and then converts the amount of delay into a register address in the decoder. The generators supply V'_{bp} and V'_{bn} for PMOS and NMOS bodies, respectively, to keep the delay constant by changing the V_{thp} and V_{thn} , respectively. If the speed of the delay line changes due to variation, the comparator output changes, and the generated body bias is modified. The junction leakage and GIDL current determine the maximum reverse-bias voltage, which was set to 1.5V , while the forward biased was limited to 0.5V to reduce the subthreshold leakage. In addition, FBB mitigates SCE, and reduces the sensitivity to channel length variation, as was shown earlier in Fig. 2.13.

This technique is efficient in dealing with D2D variations, however, it cannot mitigate WID variations effectively because this technique supplies the same body bias to the entire chip, while WID variations affect different parts of the chip differently.

A similar technique was presented in [101], where again forward and RBB were used to improve performance and decrease leakage, respectively. This adaptive body bias (ABB) allows each die to have the optimal threshold voltage which maximizes the die frequency subject to power constraint. A critical path emulator containing key circuit elements of a process critical path are used to model the effect of body bias on the frequency and leakage of the processor.

This study used multiple delay sensors distributed on the die to get an average body bias that accounts for WID variations [101]. With no body bias, only 50% of the dies meet the performance and power requirements, mainly in the lowest frequency bin. ABB using only one delay sensor reduced the frequency variation σ/μ from 4 to 1%, however, a large number of dies still failed to meet leakage constraint.

Fig. 2.41 Measured leakage versus frequency scatter for 62 dies in a 150nm technology ($V_{DD} = 1.1V$, $T=110C$), which shows the tighter distributions after utilizing ABB and WID-ABB. In the *lower figure*, the percentage of accepted dies at a certain frequency bin are shown [101]



Using multiple sensors for ABB, the frequency variation reduced to 0.69 % and all dies met the leakage constraint with 32 % in the highest frequency bin.

While the ABB scheme with several sensors considers WID variations in determining the optimum bias combination per die, it still does not completely compensate for these variations, since only a single bias combination is used per die. By allowing different blocks to have their own body bias control (WID-ABB), the effectiveness of the ABB system improves [101]. Figure 2.41 shows the results after using WID-ABB technique. More improvements can be achieved when ABB is combined with adaptive supply voltage V_{DD} [102] However, this improvement comes at the cost of additional area, design complexity, and cost.

Another body biasing technique has been proposed to reduce random fluctuations [103]. In this technique, FBB is applied to the logic circuit blocks, using a body bias generation circuit shown in Fig. 2.42. A current source determines the substrate potential by forward biasing the junction diode. The current source limits the maximum currents that the forward diodes can conduct, and the body potential is self-adjusted by the diode current. Under this self-adjusted FBB (SA-FBB) condition, $\sigma_{V_{th}}$ decreases by 35 %. Interestingly, the improvement achieved using SA-FBB was larger than the improvement using the conventional FBB technique. This improvement may be due to the fact that SA-FBB enables the body bias to become more forward biased compared to conventional FBB because the body voltage is set by

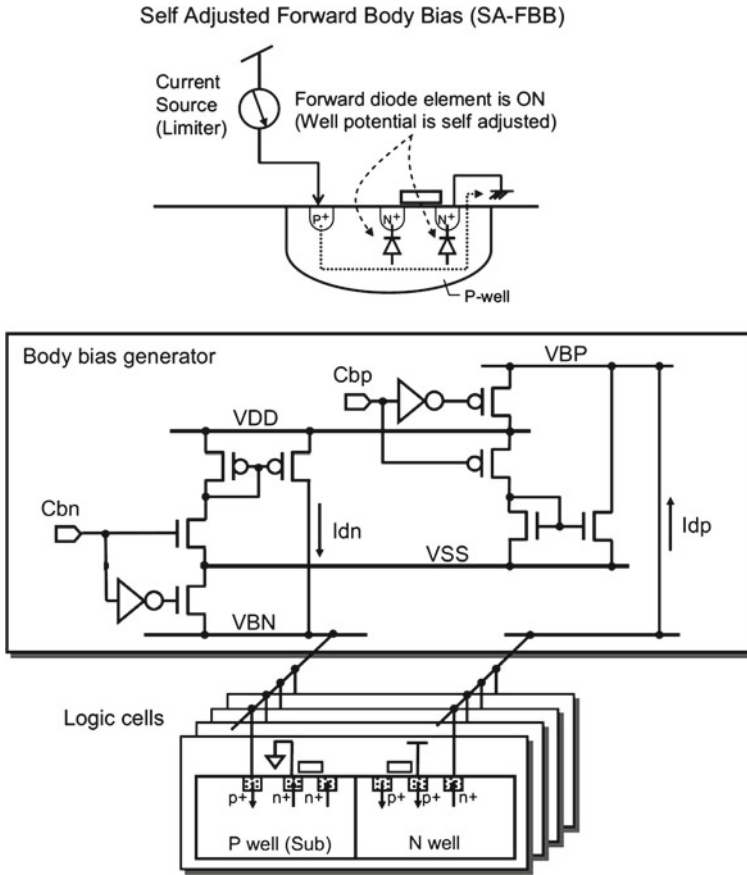


Fig. 2.42 Schematic showing the implementation of self-adjusted forward body bias (SA-FBB) circuit technique [103]

the diode current current used. In contrast, in conventional FBB, the maximum FBB is constrained by stability requirements for the substrate and preventing latch-up.

In addition to voltage and body bias control, programmable sizing has been proposed to improve the robustness of logic circuits. Dynamic circuits are usually used for high-performance gates such as high-speed register files [104]. To prevent the dynamic node from floating and hold it to V_{DD} when none of the pull-downs are evaluated, dynamic circuits use keeper devices [63]. In previous technologies, a small keeper was sufficient to hold the dynamic node to V_{DD} . As technology scales, stronger keepers are required to prevent the dynamic node from collapsing under increasing pull-down leakage levels. In addition, due to the increase in leakage variations, the keeper should be sized up which reduces the the speed of dynamic circuits and limits its advantage over static CMOS [105].

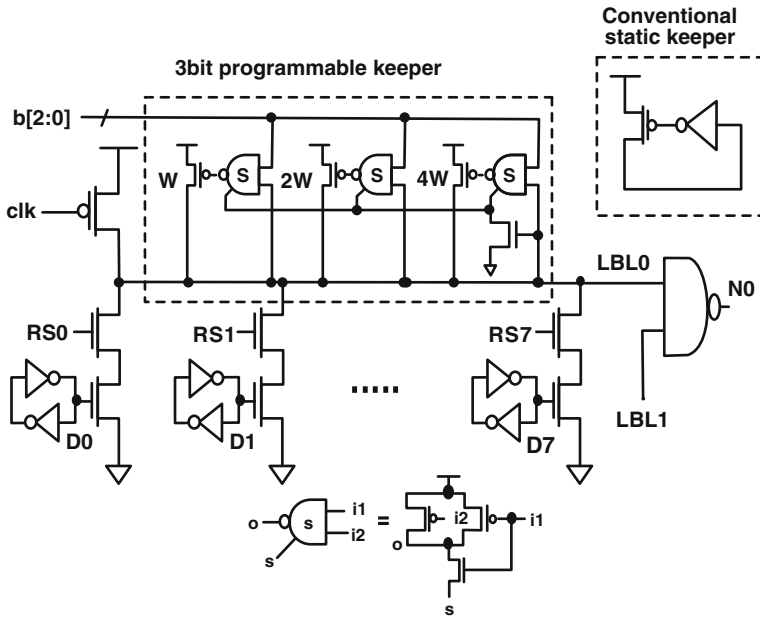


Fig. 2.43 Register file with programmable keeper to compensate for process variations impact on leakage [104]

A variation-tolerant compensation technique for dynamic circuits using programmable sizing is shown in Fig. 2.43 [104, 105]. The technique reduces the number of failing dies by 5X compared to conventional designs. An on-chip leakage sensing circuit measures leakage current, and automatically selects the optimal keeper width. In Chap. 3, we will discuss circuit techniques used to mitigate variability in SRAM.

2.7.2 Architecture

Early studies that related variability to architecture were by Bowman et al. [22, 106, 107], which presented a statistical predictive model for the distribution of the maximum operating frequency (FMAX) for a chip in the presence of process variations. The model provides insight into the impact of different components of variations on the distribution of FMAX. The WID delay distribution depends on the total number of independent critical paths for the entire chip N_{cp} . For a larger number of critical paths, the mean value of the maximum critical path delay increases as shown in Fig. 2.44. As the number of critical paths increases, the probability that one of them will be strongly affected by process variations is higher, further increasing the mean of critical path delay. On the other hand, the standard deviation (or delay spread)

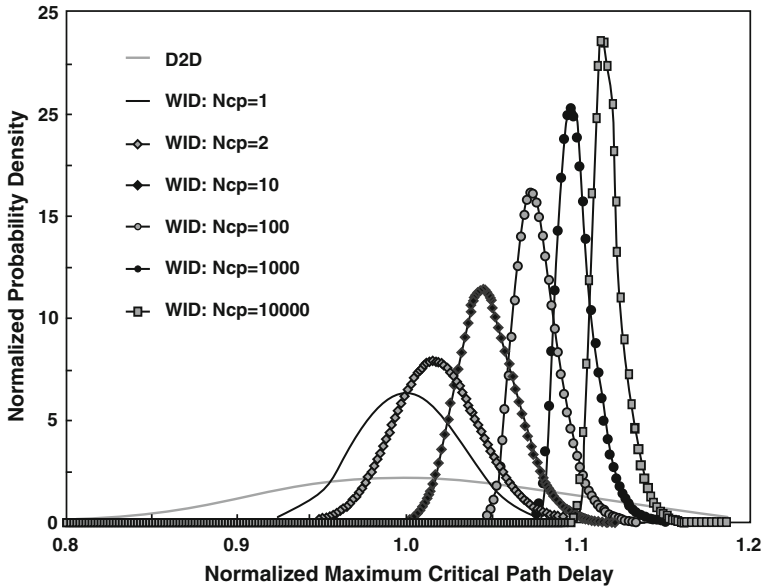


Fig. 2.44 The distribution of maximum critical path delay for different numbers of independent critical paths N_{cp} . As N_{cp} increases, the mean of maximum critical path delay increases [106]

decreases with larger N_{cp} , thus making the delay spread determined mainly by D2D variations. This work revealed that WID variations directly impact the mean of the maximum frequency, while D2D fluctuations impact the variance.

Another factor that affects the delay distribution is the logic depth per critical path. The impact of logic depth on delay distribution depends on whether WID variations are random or systematic. Random WID variations have an averaging effect on the overall critical path distribution (reduces the relative delay variation σ/μ), while systematic WID variations affect all the gates on the path, and thus, increase delay spread [106, 108, 109].

Other variation-tolerant study at the architectural level showed a statistical methodology for pipeline delay analysis [110]. This study emphasized the importance of logic depth in delay variability, and showed that changing logic depth and delay imbalance between stage delays can improve the yield of a pipeline, and showed that the push for high clock speeds using deep pipelining decreases logic depth and increases delay variability [110].

Studies on the impact of variations on low power parallel processing systems show that neglecting WID variation would underestimate the optimum supply voltage that minimizes power consumption for the system [111]. The number of parallel blocks required to meet throughput requirements increases significantly with the increase in WID process variations. As a consequence, the optimum supply voltage that provides the lowest power increases, and therefore, the targeted power parallelism becomes

less effective in decreasing power consumptions. An investigation on the impact of parameter variations on multi-core chips showed that WID variation are important for core-to-core granularity, rather than at unit-to-unit granularity [112, 113]. In Chap. 4, we will look at variation-tolerant memory architectures.

2.7.3 Statistical Timing Analysis

Considerable research has been done in the area of CAD tools that are variation-aware. One of the most researched topics in this area is statistical static timing analysis (SSTA) as compared to the well-known static timing analysis (STA) tools [13, 15, 114, 115].

The goal of timing verification is to ensure a chip will operate at a frequency with high yield under the full range of operation conditions. In timing verification, speed (setup time) and functional (hold time) are usually checked to verify that the design will meet the maximum target frequency, as well as provide correct functionality [14].

STA has been used in performance verification for the past two decades. Traditionally, process variations have been addressed in STA using corner-based analysis, where all the gates are assumed to operate at worst, typical, or best-case conditions [13]. This technique is efficient when dealing with D2D (inter-die) variation. However, since WID variation has become a substantial portion of the overall variability, corner-based STA can produce inaccurate predictions that yield low-performance designs, which has motivated the development of SSTA [13].

In SSTA, the circuit delay is considered a random variable and SSTA computes the probability density function (pdf) of the delay at a certain path [13]. The arrival times also become random variables, and therefore, the addition and maximum operations of STA are replaced by convolution and statistical maximum, respectively [13]. Much of the work on SSTA, however, has been in the area of finding efficient algorithms to perform these functions [13]. While SSTA is more appropriate in dealing with WID variations, and can give accurate results without going through the lengthy Monte Carlo simulations, the usefulness of SSTA for timing analysis has been questioned [116]. In Chap. 5, we will look at statistical CAD techniques used in SRAM design.

2.8 Summary

SRAM bitcells continue to scale by 50 % in each technology node, while the density of embedded SRAM increases significantly. At the same time, variability is worsening with technology scaling due to the increase in device variations such as RDF, LER, RTN, and device degradation (NBTI and PBTI). These types of variations have strong impact on the SRAM operation, and increase in the probability of failure in read access, read disturb, write, and hold failures. The increase in variations poses

a huge challenge for SRAM design, and determines the lowest operating voltage of the memory V_{\min} . In the subsequent chapters, circuit and architecture techniques used to mitigate SRAM variability and statistical CAD used in SRAM design will be discussed.

References

1. Y. Zorian, Embedded memory test and repair: infrastructure IP for SOC yield, in *Proceedings the International Test Conference (ITC)*, 2002, pp. 340–349
2. R. Kuppuswamy, S. Sawant, S. Balasubramanian, P. Kaushik, N. Natarajan, J. Gilbert, Over one million TPCC with a 45nm 6-core Xeon® CPU, in *IEEE International Solid-State Circuits Conference—Digest of Technical Papers, ISSCC 2009*, Febraury 2009, pp. 70–71, 71a
3. H. Pilo, IEDM SRAM short course, (2006)
4. H. Yamauchi, Embedded SRAM circuit design technologies for a 45 nm and beyond, in: *ASICON '07: 7th International Conference on ASIC*, 22–25 October 2007, pp. 1028–1033
5. K. Agarwal, S. Nassif, Statistical analysis of SRAM cell stability, in *DAC '06: Proceedings of the 43rd Annual Conference on Design Automation*, 2006, pp. 57–62
6. A. Bhavnagarwala, S. Kosonocky, C. Radens, K. Stawiasz, R. Mann, Q. Ye, K. Chin, Fluctuation limits and scaling opportunities for CMOS SRAM cells, in *Proceedings of the International Electron Devices Meeting (IEDM)*, 2005, pp. 659–662
7. H. Pilo, C. Barwin, G. Bracer, C. Browning, S. Lamphier, F. Towler, An SRAM design in 65-nm technology node featuring read and write-assist circuits to expand operating voltage. *IEEE J. Solid-State Circuits* **42**(4), 813–819 (2007)
8. M. Yamaoka, N. Maeda, Y. Shinozaki, Y. Shimazaki, K. Nii, S. Shimada, K. Yanagisawa, T. Kawahara, 90-nm process-variation adaptive embedded SRAM modules with power-line-floating write technique. *IEEE J. Solid-State Circuits* **41**(3), 705–711 (2006)
9. S. Mukhopadhyay, H. Mahmoodi, K. Roy, Statistical design and optimization of SRAM cell for yield enhancement, in *Proceedings of International Conference on Computer Aided Design*, 2004, pp. 10–13
10. The International Technology Roadmap for Semiconductors (ITRS). <http://public.itrs.net>
11. M. Sinangil, H. Mair, A. Chandrakasan, A 28 nm high-density 6T SRAM with optimized peripheral-assist circuits for operation down to 0.6 V, in *IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, Febraury 2011, pp. 260–262
12. B. Wong, A. Mittal, Y. Cao, G.W. Starr, *Nano-CMOS Circuit and Physical Design* (Wiley-Interscience, New York, 2004)
13. S. Sapatnekar, *Timing* (Springer, Boston, 2004)
14. A. Chandrakasan, W.J. Bowhill, F. Fox, *Design of High-Performance Microprocessor Circuits* (Wiley-IEEE Press, Piscataway, 2000)
15. A. Srivastava, D. Sylvester, D. Blaauw, *Statistical Analysis and Optimization for VLSI: Timing and Power (Series on Integrated Circuits and Systems)* (Springer, Boston, 2005)
16. Y. Cheng, C. Hu, *MOSFET Modeling and BSIM User Guide* (Kluwer Academic Publishers, Boston, 1999)
17. J. Tschanz, K. Bowman, V. De, Variation-tolerant circuits: circuit solutions and techniques, in *DAC '05: Proceedings of the 42nd Annual Conference on Design automation*, 2005, pp. 762–763
18. S. Borkar, T. Karnik, S. Narendra, J. Tschanz, A. Keshavarzi, V. De, Parameter variations and impact on circuits and microarchitecture, in *DAC '03: Proceedings of the 40th conference on Design automation*, 2003, pp. 338–342

19. T. Karnik, S. Borkar, V. De, Sub-90 nm technologies: challenges and opportunities for CAD, in *ICCAD '02: Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, 2002, pp. 203–206
20. S. Borkar, T. Karnik, V. De, Design and reliability challenges in nanometer technologies, in *DAC '04: Proceedings of the 41st Annual Conference on Design Automation*, 2004, pp. 75–75
21. A. Keshavarzi, G. Schrom, S. Tang, S. Ma, K. Bowman, S. Tyagi, K. Zhang, T. Linton, N. Hakim, S. Duvall, J. Brews, V. De, Measurements and modeling of intrinsic fluctuations in MOSFET threshold voltage, in *ISLPED '05: Proceedings of the 2005 International Symposium on Low Power Electronics and Design*, 2005, pp. 26–29
22. K. Bowman, S. Duvall, J. Meindl, Impact of die-to-die and within-die parameter fluctuations on the maximum clock frequency distribution for gigascale integration. *IEEE J. Solid-State Circuits* **37**(2), 183–190 (2002)
23. H. Masuda, S. Ohkawa, A. Kurokawa, M. Aoki, Challenge: variability characterization and modeling for 65- to 90-nm processes, in *Proceedings of IEEE Custom Integrated Circuits Conference*, 2005, pp. 593–599
24. J.A. Croon, W. Sansen, H.E. Maes, *Matching Properties of Deep Sub-Micron MOS Transistors* (Springer, New York, 2005)
25. Y. Taur, T.H. Ning, *Fundamentals of Modern VLSI Devices* (Cambridge University Press, New York, 1998)
26. T. Mizuno, J. Okumura, A. Toriumi, Experimental study of threshold voltage fluctuation due to statistical variation of channel dopant number in MOSFET's. *IEEE Trans. Electron Devices* **41**(11), 2216–2221 (1994)
27. K. Takeuchi, T. Tatsumi, A. Furukawa, Channel engineering for the reduction of random-dopant-placement-induced threshold voltage fluctuations, in *Proceedings of the International Electron Devices Meeting (IEDM)*, 1996, pp. 841–844
28. A. Asenov, A. Brown, J. Davies, S. Kaya, G. Slavcheva, Simulation of intrinsic parameter fluctuations in decananometer and nanometer-scale MOSFETs. *IEEE Trans. Electron Devices* **50**(9), 1837–1852 (2003)
29. K. Sonoda, K. Ishikawa, T. Eimori, O. Tsuchiya, Discrete dopant effects on statistical variation of random telegraph signal magnitude. *IEEE Trans. Electron Devices*. **54**(8), 1918–1925 (2007)
30. M. Miyamura, T. Fukai, T. Ikezawa, R. Ueno, K. Takeuchi, M. Hane, SRAM critical yield evaluation based on comprehensive physical/statistical modeling, considering anomalous non-gaussian intrinsic transistor fluctuations, *Proceedings of IEEE Symposium on VLSI Technology*, June 2007, pp. 22–23
31. M. Pelgrom, A. Duinmaijer, A. Welbers, Matching properties of MOS transistors. *IEEE J. Solid-State Circuits* **24**(5), 1433–1439 (1989)
32. D. Frank, R. Dennard, E. Nowak, P. Solomon, Y. Taur, H.S. Wong, Device scaling limits of Si MOSFETs and their application dependencies. *Proc. IEEE* **89**(3), 259–288 (2001)
33. Y. Ye, F. Liu, M. Chen, S. Nassif, Y. Cao, Statistical modeling and simulation of threshold variation under random dopant fluctuations and line-edge roughness, *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* **19**(6), 987–996 (2011)
34. B. Razavi, *Design of Analog CMOS Integrated Circuits* (McGraw-Hill, New York, 2000)
35. T.-C. Chen, Where is CMOS going: trendy hype versus real technology, in *Proceedings of the International Solid-State Circuits Conference ISSCC*, 2006, pp. 22–28
36. K. Kuhn, CMOS transistor scaling past 32 nm and implications on variation, in *Advanced Semiconductor Manufacturing Conference (ASMC), IEEE/SEMI*, July 2010, pp. 241–246
37. J.-T. Kong, CAD for nanometer silicon design challenges and success. *IEEE Trans. Very Large Scale Integr. Syst.* **12**(11), 1132–1147 (2004)
38. P. Friedberg, Y. Cao, J. Cain, R. Wang, J. Rabaey, C. Spanos, Modeling within-die spatial correlation effects for process-design co-optimization, in *ISQED '05: Proceedings of the Sixth International Symposium on Quality of Electronic Design*, 2005, pp. 516–521
39. C. Wu, Y. Leung, C. Chang, M. Tsai, H. Huang, D. Lin, Y. Sheu, C. Hsieh, W. Liang, L. Han, W. Chen, S. Chang, S. Wu, S. Lin, H. Lin, C. Wang, P. Wang, T. Lee, C. Fu, C. Chang,

- S. Chen, S. Jang, S. Shue, H. Lin, Y. See, Y. Mii, C. Diaz, B. Lin, M. Liang, Y. Sun, A 90-nm CMOS device technology with high-speed, general-purpose, and low-leakage transistors for system on chip applications, in *Proceedings of the International Electron Devices Meeting (IEDM)*, 2002, pp. 65–68
40. K. Takeuchi, T. Nagumo, K. Takeda, S. Asayama, S. Yokogawa, K. Imai, Y. Hayashi, Direct observation of RTN-induced SRAM failure by accelerated testing and its application to product reliability assessment, in *Symposium on VLSI Technology (VLSIT)*, June 2010, pp. 189–190
 41. N. Tega, H. Miki, F. Pagette, D. Frank, A. Ray, M. Rooks, W. Haensch, K. Torii, Increasing threshold voltage variation due to random telegraph noise in FETs as gate lengths scale to 20 nm, in *Symposium on VLSI Technology*, June 2009, pp. 50–51
 42. N. Tega, H. Miki, M. Yamaoka, H. Kume, T. Mine, T. Ishida, Y. Mori, R. Yamada, K. Torii, Impact of threshold voltage fluctuation due to random telegraph noise on scaled-down SRAM, in *IEEE International Reliability Physics Symposium IRPS 2008*, 27-May 1 2008, pp. 541–546
 43. S. Bhardwaj, W. Wang, R. Vattikonda, Y. Cao, S. Vrudhula, Predictive modeling of the NBTI effect for reliable design, in *Custom Integrated Circuits Conference CICC '06, IEEE*, September 2006, pp. 189–192
 44. J. Lin, A. Oates, H. Tseng, Y. Liao, T. Chung, K. Huang, P. Tong, S. Yau, Y. Wang, Prediction and control of NBTI induced SRAM Vccmin drift, in *Proceedings of the International Electron Devices Meeting (IEDM)*, 2006
 45. S. Chakravarthi, A. Krishnan, V. Reddy, C. Machala, S. Krishnan, A comprehensive framework for predictive modeling of negative bias temperature instability, in *Proceedings of the 42nd Annual IEEE International Reliability Physics Symposium*, April 2004, pp. 273–282
 46. S. Drapatz, K. Hofmann, G. Georgakos, D. Schmitt-Landsiedel, Impact of fast-recovering NBTI degradation on stability of large-scale SRAM arrays, in *Proceedings of the European Solid-State Device Research Conference (ESSDERC)*, September 2010, pp. 146–149
 47. J. Lin, A. Oates, C. Yu, Time dependent Vccmin degradation of SRAM fabricated with High-k gate dielectrics, in *proceedings of the IEEE International 45th Annual Reliability Physics Symposium*, April 2007, pp. 439–444
 48. P. Kinget, Device mismatch and tradeoffs in the design of analog circuits. *IEEE J. Solid-State Circuits* **40**(6), 1212–1224 (2005)
 49. E. Karl, Y. Wang, Y.-G. Ng, Z. Guo, F. Hamzaoglu, U. Bhattacharya, K. Zhang, K. Mistry, M. Bohr, A 4.6 GHz 162 Mb SRAM design in 22 nm tri-gate CMOS technology with integrated active VMIN-enhancing assist circuitry, in *IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, February 2012, pp. 230–232
 50. S. Damaraju, V. George, S. Jahagirdar, T. Khondker, R. Milstrey, S. Sarkar, S. Siers, I. Stoloro, A. Subbiah, A 22 nm IA multi-CPU and GPU system-on-chip, in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, *IEEE International*, February 2012, pp. 56–57
 51. K. Kuhn, Reducing variation in advanced logic technologies: Approaches to process and design for manufacturability of nanoscale CMOS, in *IEEE International Electron Devices Meeting IEDM 2007*, December 2007, pp. 471–474
 52. Y. Liu, K. Endo, S. O'uchi, T. Kamei, J. Tsukada, H. Yamauchi, Y. Ishikawa, T. Hayashida, K. Sakamoto, T. Matsukawa, A. Ogura, and M. Masahara, On the gate-stack origin threshold voltage variability in scaled FinFETs and multi-FinFETs, in *Symposium on VLSI Technology (VLSIT)*, June 2010, pp. 101–102
 53. M. Khellah, D. Khalil, D. Somasekhar, Y. Ismail, T. Karnik, V. De, Effect of power supply noise on SRAM dynamic stability, *Proceedings of IEEE Symposium on VLSI Circuits*, June 2007, pp. 76–77
 54. R. Heald, P. Wang, Variability in sub-100 nm SRAM designs, in *Proceedings of International Conference on Computer Aided Design*, 2004, pp. 347–352
 55. M. Yamaoka, N. Maeda, Y. Shinozaki, Y. Shimazaki, K. Nii, S. Shimada, K. Yanagisawa, T. Kawahara, Low-power embedded SRAM modules with expanded margins for writing, in *Proceedings of the International Solid-State Circuits Conference ISSCC*, 2005, vol. 1, pp. 480–611

56. M. Yabuuchi, K. Nii, Y. Tsukamoto, S. Ohbayashi, S. Imaoka, H. Makino, Y. Yamagami, S. Ishikura, T. Terano, T. Oashi, K. Hashimoto, A. Sebe, G. Okazaki, K. Satomi, H. Akamatsu, H. Shinohara, A 45 nm low-standby-power embedded SRAM with improved immunity against process and temperature variations, in *Proceedings of the International Solid-State Circuits Conference ISSCC*, 11–15 February 2007, pp. 326–606
57. H. Qin, Y. Cao, D. Markovic, A. Vladimirescu, J. Rabaey, SRAM leakage suppression by minimizing standby supply voltage, in *Proceedings of the International Symposium on Quality of Electronic Design ISQED*, 2004, pp. 55–60
58. J. Wang, A. Singhee, R. Rutenbar, B. Calhoun, Statistical modeling for the minimum standby supply voltage of a full SRAM array, in *33rd European Solid State Circuits Conference ESSCIRC*, September 2007, pp. 400–403
59. M.H. Abu-Rahma, K. Chowdhury, J. Wang, Z. Chen, S.S. Yoon, M. Anis, A methodology for statistical estimation of read access yield in SRAMs, in *DAC '08: Proceedings of the 45th Conference on Design Automation*, 2008, pp. 205–210
60. M. Yamaoka, T. Kawahara, Operating-margin-improved SRAM with column-at-a-time body-bias control technique, in *33rd European Solid State Circuits Conference ESSCIRC*, 11–13 September 2007, pp. 396–399
61. E. Seevinck, F. List, J. Lohstroh, Static-noise margin analysis of MOS SRAM cells. *IEEE J. Solid-State Circuits* **22**(5), 748–754 (1987)
62. K. Takeda, H. Ikeda, Y. Hagihara, M. Nomura, H. Kobatake, Redefinition of write margin for next-generation SRAM and write-margin monitoring circuit, in *Proceedings of the International Solid-State Circuits Conference ISSCC*, 2006, pp. 2602–2611
63. J.M. Rabaey, A. Chandrakasan, B. Nikolic, *Digital Integrated Circuits*, 2nd Edn. (Prentice Hall, Englewood Cliffs, 2002)
64. K. Aadithya, S. Venogopalan, A. Demir, J. Roychowdhury, Mustard: a coupled, stochastic/deterministic, discrete/continuous technique for predicting the impact of random telegraph noise on SRAMs and DRAMs, in *Design Automation Conference (DAC)*, 2011 48th ACM/EDAC/IEEE, June 2011, pp. 292–297
65. F. shi Lai, C.-F. Lee, On-chip voltage down converter to improve SRAM read/write margin and static power for sub-nano CMOS technology, *IEEE J. Solid-State Circuits* **42**(9), 2061–2070 (2007)
66. A. Carlson, Mechanism of increase in SRAM Vmin due to negative-bias temperature instability, *IEEE Trans. Device Mater. Reliab.* **7**(3), 473–478 (2007)
67. A. Krishnan, V. Reddy, D. Aldrich, J. Raval, K. Christensen, J. Rosal, C. O'Brien, R. Khamankar, A. Marshall, W.-K. Loh, R. McKee, S. Krishnan, SRAM cell static noise margin and VMIN sensitivity to transistor degradation, in *Proceedings of the International Electron Devices Meeting (IEDM)*, 2006
68. G. La Rosa, W.L. Ng, S. Rauch, R. Wong, J. Sudijono, Impact of NBTI induced statistical variation to SRAM cell stability, in *Proceedings of the 44th Annual IEEE International Reliability Physics Symposium*, March 2006, pp. 274–282
69. K. Kang, H. Kufluoglu, K. Roy, M. Ashraful Alam, Impact of negative-bias temperature instability in nanoscale SRAM array: Modeling and analysis. *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.* **26**(10), 1770–1781 (2007)
70. M. Ball, J. Rosal, R. McKee, W. Loh, T. Houston, R. Garcia, J. Raval, D. Li, R. Hollingsworth, R. Gury, R. Eklund, J. Vaccani, B. Castellano, F. Piacibello, S. Ashburn, A. Tsao, A. Krishnan, J. Ondrusek, T. Anderson, A screening methodology for VMIN drift in SRAM arrays with application to sub-65 nm nodes, in *International Electron Devices Meeting, IEDM '06*, December 2006, pp. 1–4
71. M. Agostinelli, J. Hicks, J. Xu, B. Woolery, K. Mistry, K. Zhang, S. Jacobs, J. Jopling, W. Yang, B. Lee, T. Raz, M. Mehalel, P. Kolar, Y. Wang, J. Sandford, D. Pivin, C. Peterson, M. DiBattista, S. Pae, M. Jones, S. Johnson, G. Subramanian, Erratic fluctuations of SRAM cache vmin at the 90 nm process technology node, in *Electron Devices Meeting IEDM Technical Digest. IEEE International*, December 2005, pp. 655–658

72. M. Tanizawa, S. Ohbayashi, T. Okagaki, K. Sonoda, K. Eikyu, Y. Hirano, K. Ishikawa, O. Tsuchiya, Y. Inoue, Application of a statistical compact model for random telegraph noise to scaled-SRAM Vmin analysis, in *Symposium on VLSI Technology (VLSIT)*, June 2010, pp. 95–96
73. S. O. Toh, T.-J. K. Liu, B. Nikolic, Impact of random telegraph signaling noise on SRAM stability, in *Symposium on VLSI Technology (VLSIT)*, June 2011, pp. 204–205
74. M. Yamaoka, H. Miki, A. Bansal, S. Wu, D. Frank, E. Leobandung, K. Torii, Evaluation methodology for random telegraph noise effects in SRAM arrays, in *IEEE International Electron Devices Meeting (IEDM)*, December 2011, pp. 32.2.1–32.2.4
75. S.O. Toh, Y. Tsukamoto, Z. Guo, L. Jones, T.-J.K. Liu, B. Nikolic, Impact of random telegraph signals on Vmin in 45 nm SRAM, in *IEEE International Electron Devices Meeting (IEDM)*, December 2009, pp. 1–4
76. K. Takeuchi, T. Nagumo, T. Hase, Comprehensive SRAM design methodology for RTN reliability, in *Symposium on VLSI Technology (VLSIT)*, June 2011, pp. 130–131
77. Y. Tsukamoto, S. O. Toh, C. Shin, A. Mairena, T.-J. K. Liu, B. Nikolić, Analysis of the relationship between random telegraph signal and negative bias temperature instability, in *IEEE International Reliability Physics Symposium (IRPS)*, May 2010, pp. 1117–1121
78. R. Baumann, The impact of technology scaling on soft error rate performance and limits to the efficacy of error correction, in *Proceedings of the International Electron Devices Meeting (IEDM)*, 2002, pp. 329–332
79. R. Baumann, Soft errors in advanced computer systems, *IEEE Des. Test of Comput.* **22**(3) 258–266 (2005)
80. E. Cannon, D. Reinhardt, M. Gordon, P. Makowskyj, SRAM SER in 90, 130 and 180 nm bulk and SOI technologies, in *Proceedings of the 42nd IEEE International Reliability Physics Symposium*, 2004, pp. 300–304
81. A. Balasubramanian, P. Fleming, B. Bhuvu, A. Sternberg, L. Massengill, Implications of dopant-fluctuation-induced v_t variations on the radiation hardness of deep submicrometer CMOS SRAMs. *IEEE Trans. Device and Mater. Reliab.* **8**(1), 135–144 (2008)
82. E. Cannon, A. KleinOowski, R. Kanj, D. Reinhardt, R. Joshi, The impact of aging effects and manufacturing variation on SRAM soft-error rate, *IEEE Trans. Device Mater. Reliab.* **8**(1), 145–152 (2008)
83. S.S. Mukherjee, J. Emer, S.K. Reinhardt, The soft error problem: an architectural perspective, in *HPCA '05: Proceedings of the 11th International Symposium on High-Performance Computer Architecture*, 2005, pp. 243–247
84. E. Ibe, H. Taniguchi, Y. Yahagi, K.-i. Shimbo, and T. Toba, Impact of scaling on neutron-induced soft error in SRAMs from a 250 nm to a 22 nm design rule, *IEEE Trans. Electron Devices* **57**(7), 1527–1538 (2010)
85. A. Dixit, A. Wood, The impact of new technology on soft error rates, in *Reliability Physics Symposium (IRPS)*, 2011 *IEEE International*, April 2011, pp. 5B.4.1–5B.4.7
86. N.H. Weste, D. Harris, *CMOS VLSI Design : A Circuits and Systems Perspective*, 3rd Edn. (Addison Wesley, Boston, 2004)
87. K. Itoh, M. Horiguchi, M. Yamaoka, Low-voltage limitations of memory-rich nano-scale CMOS LSIs, in *37th European Solid State Device Research Conference, ESSDERC 2007*, September 2007, pp. 68–75
88. R. Rajsuman, Design and test of large embedded memories: an overview. *IEEE Des. Test Comput.* **18**(3), 16–27 (2001)
89. Y. Zorian, S. Shoukourian, Embedded-memory test and repair: infrastructure IP for SoC yield. *IEEE Des. Test Comput.* **20**(3), 58–66 (2003)
90. T. Chen, V.-K. Kim, M. Tegethoff, IC yield estimation at early stages of the design cycle. *Microelectron. J.* **30**(8), 725–732 (1999)
91. T. Barnett, M. Grady, K. Purdy, A. Singh, Redundancy implications for early-life reliability: experimental verification of an integrated yield-reliability model, in *Proceedings of the International Test Conference*, 2002, pp. 693–699

92. K. Roy, S. Mukhopadhyay, H. Mahmoodi-Meimand, Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits. *Proc. IEEE* **91**(2), 305–327 (2003)
93. Y.-S. Lin, C.-C. Wu, C.-S. Chang, R.-P. Yang, W.-M. Chen, J.-J. Liaw, C. Diaz, Leakage scaling in deep submicron CMOS for SoC. *IEEE Trans. Electron Devices* **49**(6), 1034–1041 (2002)
94. J. Kao, S. Narendra, A. Chandrakasan, Subthreshold leakage modeling and reduction techniques, in *ICCAD '02: Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, 2002, pp. 141–148
95. S. Narendra, V. De, S. Borkar, D. Antoniadis, A. Chandrakasan, Full-chip subthreshold leakage power prediction and reduction techniques for sub-0.18 μm CMOS. *IEEE J. Solid-State Circuits* **39**(2), 501–510 (2004)
96. J.P. de Gyvez, H. Tuinhout, Threshold voltage mismatch and intra-die leakage current in digital CMOS circuits. *IEEE J. Solid-State Circuits* **39**(1), 157–168 (2004)
97. S. Narendra, V. De, S. Borkar, D. Antoniadis, A. Chandrakasan, Full-chip sub-threshold leakage power prediction model for sub-0.18 μm CMOS, in *ISLPED '02: Proceedings of the 2002 International Symposium on Low Power Electronics and Design*, 2002, pp. 19–23
98. R. Rao, A. Srivastava, D. Blaauw, D. Sylvester, Statistical analysis of subthreshold leakage current for VLSI circuits. *IEEE Trans Very Large Scale Integr. Syst.* **12**(2), 131–139 (2004)
99. M. Miyazaki, G. Ono, T. Hattori, K. Shiozawa, K. Uchiyama, K. Ishibashi, A 1000-MIPS/w microprocessor using speed adaptive threshold-voltage CMOS with forward bias, in *Proceedings of the International Solid-State Circuits Conference ISSCC*, 2000, pp. 420–421
100. M. Miyazaki, G. Ono, K. Ishibashi, A 1.2-GIPS/w microprocessor using speed-adaptive threshold-voltage CMOS with forward bias. *IEEE J. Solid-State Circuits* **37**(2), 210–217 (2002)
101. J. Tschanz, J. Kao, S. Narendra, R. Nair, D. Antoniadis, A. Chandrakasan, V. De, Adaptive body bias for reducing impacts of die-to-die and within-die parameter variations on microprocessor frequency and leakage. *IEEE J. Solid-State Circuits* **37**(11), 1396–1402 (2002)
102. J. Tschanz, S. Narendra, R. Nair, V. De, Effectiveness of adaptive supply voltage and body bias for reducing impact of parameter variations in low power and high performance microprocessors. *IEEE J. Solid-State Circuits* **38**(5), 826–829 (2003)
103. Y. Komatsu, K. Ishibashi, M. Yamamoto, T. Tsukada, K. Shimazaki, M. Fukazawa, M. Nagata, Substrate-noise and random-fluctuations reduction with self-adjusted forward body bias, in *Proceedings of IEEE Custom Integrated Circuits conference*, 2005, pp. 35–38
104. C. Kim, K. Roy, S. Hsu, A. Alvandpour, R. Krishnamurthy, S. Borkar, A process variation compensating technique for sub-90 nm dynamic circuits, in *Proceedings of IEEE Symposium on VLSI Circuits*, 2003, pp. 205–206
105. C. Kim, S. Hsu, R. Krishnamurthy, S. Borkar, K. Roy, Self calibrating circuit design for variation tolerant VLSI systems, in *IOLTS 2005 11th IEEE International On-Line Testing Symposium*, 2005, pp. 100–105
106. K. Bowman, S. Duvall, J. Meindl, Impact of die-to-die and within-die parameter fluctuations on the maximum clock frequency distribution, in *Proceedings of the International Solid-State Circuits Conference ISSCC*, 2001, pp. 278–279
107. K. Bowman, J. Meindl, Impact of within-die parameter fluctuations on future maximum clock frequency distributions, in *Proceedings of IEEE Custom Integrated Circuits Conference*, 2001, pp. 229–232
108. D. Marculescu, E. Talpes, Variability and energy awareness: a microarchitecture-level perspective, in *DAC '05: Proceedings of the 42nd Annual Conference on Design Automation*, 2005, pp. 11–16
109. D. Marculescu, E. Talpes, Energy awareness and uncertainty in microarchitecture-level design. *IEEE Micro*. **25**(5), 64–76 (2005)
110. A. Datta, S. Bhunia, S. Mukhopadhyay, N. Banerjee, K. Roy, Statistical modeling of pipeline delay and design of pipeline under process variation to enhance yield in sub-100 nm technologies, in *DATE '05: Proceedings of the Conference on Design, Automation and Test in Europe*, 2005, pp. 926–931

111. N. Azizi, M.M. Khellah, V. De, F.N. Najm, Variations-aware low-power design with voltage scaling, in *DAC '05: Proceedings of the 42nd Annual Conference on Design Automation*, 2005, pp. 529–534
112. E. Humenay, D. Tarjan, K. Skadron, Impact of parameter variations on multi-core chips, in *Proceedings of the 2006 Workshop on Architectural Support for Gigascale Integration, in Conjunction with the 33rd International Symposium on Computer Architecture (ISCA)*, June 2006
113. E. Humenay, W. Huang, M.R. Stan, K. Skadron, Toward an architectural treatment of parameter variations, University of Virginia, Department of Computer Science, Technical Report. CS-2005-16, September 2005
114. S.H. Choi, B.C. Paul, K. Roy, Novel sizing algorithm for yield improvement under process variation in nanometer technology, in *DAC '04: Proceedings of the 41st Annual Conference on Design Automation*, 2004, pp. 454–459
115. A. Agarwal, K. Chopra, D. Blaauw, Statistical timing based optimization using gate sizing, in *DATE '05: Proceedings of the Conference on Design, Automation and Test in Europe*, 2005, pp. 400–405
116. F.N. Najm, On the need for statistical timing analysis, in *DAC '05: Proceedings of the 42nd Annual Conference on Design Automation*, 2005, pp. 764–765

Nanometer Variation-Tolerant SRAM
Circuits and Statistical Design for Yield

Abu Rahma, M.; Anis, M.

2013, XVI, 172 p., Hardcover

ISBN: 978-1-4614-1748-4