Chapter 2
Temperature Effects in Semiconductors

The changes in temperature described in the previous chapter affect the speed, power, and reliability of our systems. Throughout this book, we will examine all three of these metrics, though the majority of our discussion will be on how temperature affects the speed performance. In this chapter, we discuss the problem of temperature variation at the device and circuit level. In Sect. 2.1, we provide a background on the material dependences on temperature. In Sect. 2.2, the normal and reverse temperature dependence regimes are described. In Sect. 2.3, we explore how these dependences change with technology scaling and the introduction of new processing materials, such as high-κ dielectrics and metal gates.

2.1 Material Temperature Dependences

In this section we provide details about the impact of temperature on the MOSFET energy band gap, carrier density, mobility, carrier diffusion, velocity saturation, current density, threshold voltage, leakage current, interconnect resistance, and electromigration.

2.1.1 Energy Band Gap

Temperature affects the properties of electronic systems in a number of fundamental ways. The most fundamental of properties is the energy band gap, E_g, which is affected by temperature according to the Varshni equation [1]

$$E_g(T) = E_g(0) - \frac{\alpha_E T^2}{T + \beta_E} \tag{2.1}$$

where $E_g(0)$ is the band gap energy at absolute zero on the Kelvin scale in the given material, and α_E and β_E are material-specific constants. Table 2.1 [2] provides these

D. Wolpert and P. Ampadu, *Managing Temperature Effects in Nanoscale Adaptive Systems*, DOI 10.1007/978-1-4614-0748-5_2,

Table 2.1 Varshni equation constants for GaAs, Si, and Ge [2]

Material	$E_g(0)$ (eV)	α_E (eV/K)	β_E (K)
GaAs	1.519	$5.41*10^{-4}$	204
Si	1.170	$4.73*10^{-4}$	636
Ge	0.7437	$4.77*10^{-4}$	235

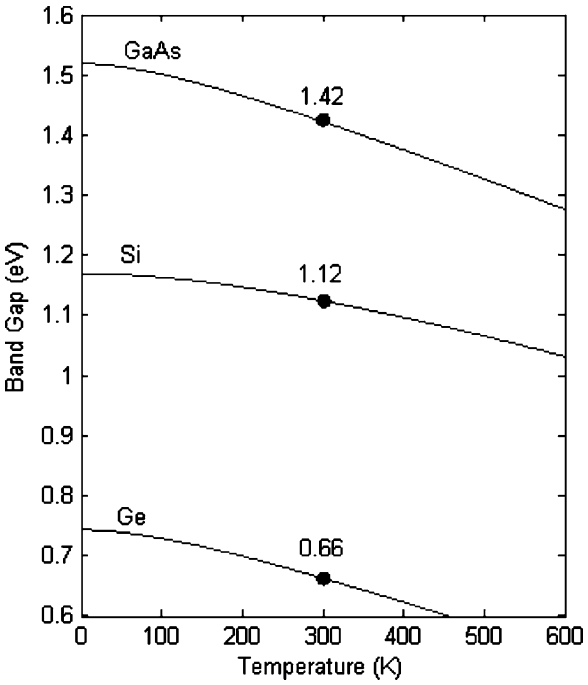

Fig. 2.1 Energy band gap temperature dependence of GaAs, Si, and Ge

constants for three material structures. Table 2.1 and (2.1) are used to generate Fig. 2.1, which shows how the band gaps of the three materials decrease as temperature increases (the labeled points are the band gap of each material at room temperature).

2.1.2 Carrier Density

Carrier densities affect electrical and thermal conductivity, and are a function of the effective density of states in the appropriate band (conduction for n-type, valence for p-type), the Fermi energy level in the material (which is a function of temperature and dopant concentrations), and the temperature as given by the following equations:

$$n = N_C e^{-\frac{E_C - E_F}{kT}} \tag{2.2}$$

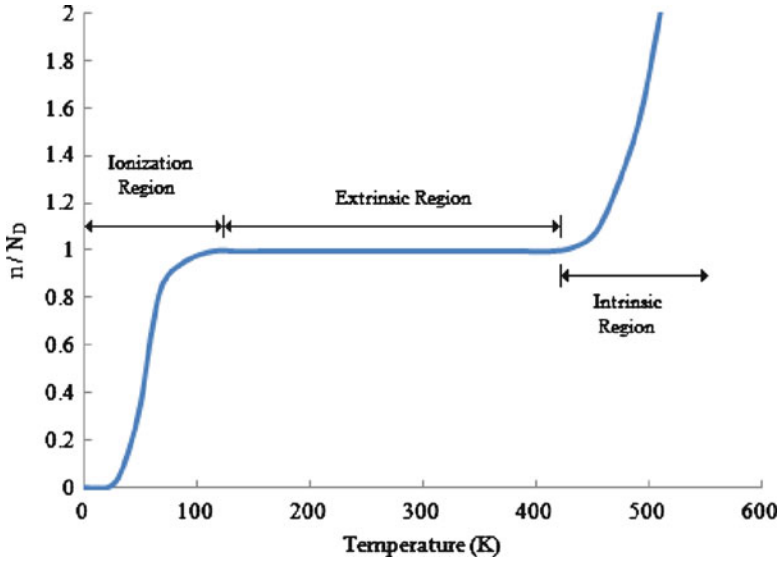

Fig. 2.2 Temperature dependence of n in a doped semiconductor

$$p = N_V e^{-\frac{E_F - E_V}{kT}} \tag{2.3}$$

where n is the electron density, p is the hole density, N_C is the density of states in the conduction band, N_V is the density of states in the valence band, E_C is the conduction band energy level, E_V is the valence band energy level, E_F is the Fermi energy level, $k = 1.38{\cdot}10^{-23}$ J/K is the Boltzmann constant, and T is temperature.

The temperature dependence of carrier density is shown in Fig. 2.2 for a doped material. In the ionization region, there is only enough latent energy in the material to push a few of the dopant carriers into the conduction band. In the extrinsic region, which is the desired region of operation, the carrier concentration is flat over a wide range of temperatures; in this region, all of the dopant carriers have been energized into the conduction band (i.e. $n \approx N_D$) and there is very little thermal generation of additional carriers. As the temperature increases, the extrinsic region turns into the intrinsic region, and the number of thermally generated carriers exceeds the number of donor carriers. The intrinsic carrier concentration in a material n_i is generally much smaller than the dopant carrier concentration at room temperature, but n_i ($=n{\cdot}p$) has a very strong temperature dependence [2]

$$n_i \propto T^{1.5} e^{-\frac{E_{g0}}{2kT}} \tag{2.4}$$

where E_{g0} is the energy band gap at $T = 0$ K. Depending upon the dopant concentration, the number of thermally generated carriers can exceed the number of dopant-generated carriers, increasing the potential for thermal variation problems.

2.1.3 Mobility

We pay particular attention to the temperature and electric field dependence of mobility, as mobility is one of the two main factors (the other is threshold voltage) resulting in the MOSFET temperature behavior shown later in this chapter. The carrier mobility, μ ($cm^2/V{\cdot}s$), describes the drift velocity of a particle in an applied electric field. Under small to moderate electric fields, $\mu = v_d/\xi$ where v_d is the drift velocity, and ξ is the electric field. MOSFET mobility has very complex temperature dependence, defined by the interplay of the following four scattering parameters: phonon scattering μ_{ph}, surface roughness scattering μ_{sr}, bulk charge Coulombic scattering μ_{cb}, and interface charge Coulombic scattering μ_{int} [3]. Each of these scattering parameters is related to the temperature of the material, T, and the effective transverse electric field in the channel, ξ_{eff}, which is approximated as [4, 5]

$$\xi_{eff} \approx \frac{(\eta Q_{inv} + Q_b)}{\varepsilon_{Si}} \approx \frac{(V_{gs} + V_T)}{6T_{ox}} \tag{2.5}$$

where η is a constant ($\eta \approx 0.4$ in PMOS devices and $\eta \approx 0.5$ in NMOS devices), Q_{inv} is the inversion layer charge density, Q_b is the substrate depletion charge density, and $\varepsilon_{Si} = 11.7$ is the relative permittivity of Silicon. This approximation is not very convenient for circuit analysis, so ξ_{eff} is also approximated in terms of the gate-source voltage V_{gs}, the threshold voltage V_T, and gate oxide thickness T_{ox}.

The Berkeley Short-Channel IGFET Model (BSIM), one of the most widely used simulation models, combines these four scattering parameters into an effective mobility, μ_{eff} [3] using Matthiessen's rule

$$\frac{1}{\mu_{eff}(T, E_{eff})} \propto \frac{1}{\mu_{ph}(T, E_{eff})} + \frac{1}{\mu_{sr}(T, E_{eff})} + \frac{1}{\mu_{cb}(T, E_{eff})} + \frac{1}{\mu_{\text{int}}(T, E_{eff})} \tag{2.6}$$

Phonon scattering refers to the potential for an electron to be scattered by a lattice vibration. As temperature increases, lattice vibrations increase and the probability of an electron being scattered by the lattice increases; thus, high temperature mobilities are limited by phonon scattering ($\mu_{ph} \propto T^{-3/2}$), causing mobility to decrease as temperature increases as shown in Fig. 2.3a. Surface roughness scattering becomes dominant when high electric fields pull electrons closer to the Si/SiO_2 surface ($\mu_{sr} \propto {\xi_{eff}}^{-2.1}$).

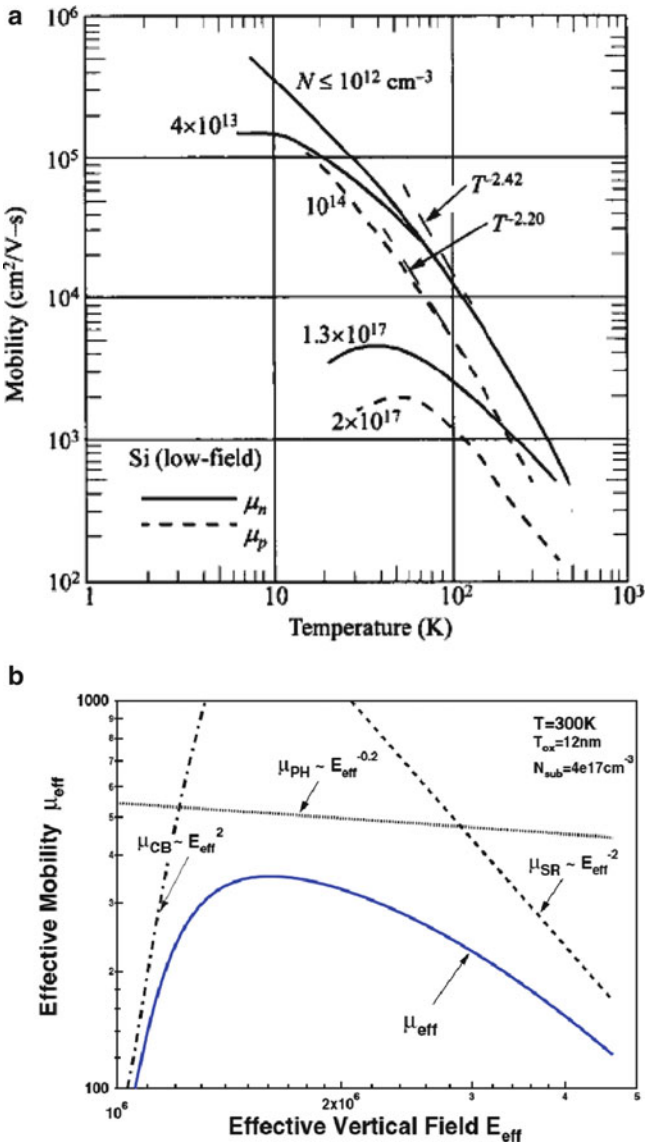

Fig. 2.3 (**a**) Temperature dependence of electron and hole mobilities in Si for different dopant concentrations [2], (**b**) Field dependence of mobility [7]

At low temperatures, electrons move more slowly, and lattice vibrations are small as well; thus, the ion impurity forces which have little impact on high-energy particles become the dominant limit to mobility. In this regime, decreasing temperature extends the amount of time electrons spend passing an impurity ion, causing mobility to decrease as temperature decreases ($\mu_{cb} \propto T$). This effect is emphasized

in the high dopant concentration curves shown in Fig. 2.3a, where mobility decreases with decreasing temperature (e.g. the $\mu_n = 1.3{\cdot}10^{17}$ dopant concentration line below ~30 K).

The electric field dependence of mobility is shown in Fig. 2.3b. In bulk Coulombic scattering, increasing ξ_{eff} increases the charge density in the channel; the associated charge screening reduces the impact of μ_{cb} ($\propto \xi_{eff}^{2}$). At low temperatures, the interface charges have two conflicting dependences. Reduced temperature reduces the carriers' thermal velocity, which increases the impact of interface charges; however, the reduced thermal velocity also reduces the screening effect [6], and this reduction in screening dominates the temperature dependence ($\mu_{int} \propto T^{-1}$). The electric field screening effect is also weakened by the reduced thermal velocity ($\mu_{\text{int}} \propto \xi_{eff}$, not ξ_{eff}^{2} as in the μ_{cb} limit). In this book, we consider devices operating in the phonon scattering limit, with temperatures >200 K; thus, mobility will decrease as temperature increases.

The temperature dependence of mobility plays a major role in temperature-aware system design, and is discussed in more detail in the next subsection. In room temperature Si, the electron mobility, μ_n, is nearly three times as large as the hole mobility, μ_p, with $\mu_n = 1{,}350$ cm^2/V·s and $\mu_p = 480$ cm^2/V·s.

2.1.4 Carrier Diffusion

Diffusion is the movement of particles from a region of high concentration to a region of low concentration. Carrier diffusion coefficients D_n and D_p (for electrons and holes, respectively) are related to mobility by the Einstein relationship

$$\frac{D}{\mu} = \frac{kT}{q} \tag{2.7}$$

Here, q is the charge on an electron ($1.6{\cdot}10^{-19}$ C), and kT/q is an important value known as the thermal voltage, ϕ_T. At room temperature (300 K), $\phi_T = 0.0259$ V. D_n and D_p in room temperature silicon are 36 and 12 cm^2/s, respectively.

2.1.5 Velocity Saturation

Although saturation velocity has been recently found to be a dominant temperature-dependent parameter, notable work had been performed in this area as far back as 1970 [8] using device lengths of 10 μm. In the BSIM4 device model, the impact of temperature on velocity saturation v_{sat} is modeled by [9]

$$v_{sat} = v_{sat0} \cdot [1 - \alpha_{v_{sat}}(T - T_0)] \tag{2.8}$$

where v_{sat0} is the saturation velocity at nominal temperature (T_0) and α_{vsat} is the saturation velocity temperature coefficient. Qualitatively, velocity saturation is the point at which increases in energy no longer cause carrier velocity to increase; instead, the additional energy is lost to phonon generation through lattice interactions.

In the results presented in this book, devices operate in the velocity saturation regime; thus, the impact of temperature on saturation velocity (increasing temperature decreases v_{sat}) is one of the most important criteria affecting the overall impact of temperature on device current, as will be shown later in this chapter.

2.1.6 Current Density

The temperature dependence of the carrier concentrations, mobilities and diffusion coefficients affect the temperature behavior of the carrier current densities, with the carrier densities defined by the following formulas [10]:

$$J_N = q\mu_n n\xi + qD_n \nabla n \tag{2.9}$$

$$J_P = q\mu_p p\xi - qD_p \nabla p \tag{2.10}$$

where J_N and J_P are the electron and hole current densities, respectively. The first term in each equation is the *drift* component of the total current, with μ_n and μ_p corresponding to the electron and hole mobilities, respectively; ξ is the electric field. The second term in each equation is the *diffusion* component of the total current, with ∇n and ∇p corresponding to the electron and hole concentration gradients (if there is no concentration gradient, there is no diffusion). The temperature dependent parameter in the second term is the diffusion coefficient. Increased temperature increases particle kinetic energy, increasing the diffusion component of total current. The drift component of the total current has two temperature dependent parameters, the mobility and the carrier density. The mobility term was shown in Fig. 2.3 to decrease as temperature increases (in the lattice vibration-limited case) while the carrier density remains nearly fixed with temperature over the extrinsic range (our intended range of operation), as indicated by Fig. 2.2. Thus, we determine that the drift component of the total current decreases as temperature increases.

The drift and diffusion currents have opposing temperature dependencies, which causes the net current change to depend on the applied electric field. In the high-field (drift-dominated) case, current decreases as temperature increases; in the low-field (diffusion-dominated) case, current increases as temperature increases. However, if the system in question is a multi-voltage system, and the system has both drift- *and* diffusion-dominated components, the impact of temperature variation may become less well-defined. The difference between a drift-dominated system and a diffusion-dominated system is defined by the threshold voltage, V_T. We will show that the temperature dependences of mobility and threshold voltage result in some very interesting device behavior.

2.1.7 Threshold Voltage

The MOSFET threshold voltage is given by [2]

$$V_T = V_{FB} + 2\phi_F + \gamma\sqrt{2\phi_F} \tag{2.11}$$

where $V_{FB} = \phi_{gs} - (Q_{ss}/C_{ox})$ is the flat band voltage, with the gate-substrate contact potential $\phi_{gs} = \phi_T \cdot \ln(N_A N_G / {n_i}^2)$, N_A and N_G are the substrate and gate doping concentrations, respectively, Q_{ss} the surface charge density, and C_{ox} the oxide capacitance; $\gamma = C_{ox}(2q\varepsilon_{Si}N_A)^{0.5}$ is a body effect parameter, with ε_{Si} the relative permittivity of Si; $\phi_F = \phi_T \cdot \ln(N_A/n_i)$ is the Fermi energy with the thermal voltage $\phi_T = kT/q$, and n_i the intrinsic carrier concentration of Si.

Of the parameters in (2.11), ϕ_{gs} and ϕ_F vary with temperature (each contains ϕ_T and n_i terms). The threshold voltage temperature dependence $\partial V_T / \partial T$ may thus be written as [11]

$$\frac{\partial V_T}{\partial T} = \frac{\partial \phi_{gs}}{\partial T} + 2\frac{\partial \phi_F}{\partial T} + \frac{\gamma}{\sqrt{2\phi_F}}\frac{\partial \phi_F}{\partial T} \tag{2.12}$$

where the temperature dependencies of ϕ_{gs} and ϕ_F are [11]

$$\frac{\partial \phi_{gs}}{\partial T} = \frac{1}{T}\left(\phi_{gs} + \frac{E_{G0}}{q} + \frac{3kT}{q}\right) \tag{2.13}$$

$$\frac{\partial \phi_F}{\partial T} = \frac{1}{T}\left[\phi_F - \left(\frac{E_{G0}}{2q} + \frac{3kT}{2q}\right)\right] \tag{2.14}$$

Filanovsky [11] used empirical parameters from a 0.35 μm CMOS technology to determine that the three terms in (2.12) are −3.1, 2.7, and −0.43 mV/K, resulting in a net threshold temperature coefficient of −0.83 mV/K. The threshold voltage in a MOSFET is commonly modeled to decrease linearly with increasing temperature; the parameter is plotted in Fig. 2.4 over a range of oxide thicknesses d and dopant concentrations N_A.

2.1.8 Leakage Current

Subthreshold leakage current I_{sub} is exponentially dependent on temperature, as shown in Fig. 2.5; a common rule of thumb is that leakage current doubles for every 10°C increase in temperature [12]. When $V_{GS} = 0$, I_{sub} may be represented by the Shockley diode model

$$I_{sub} = I_0\left(e^{\frac{V_{DS}}{\phi_T}} - 1\right) \tag{2.15}$$

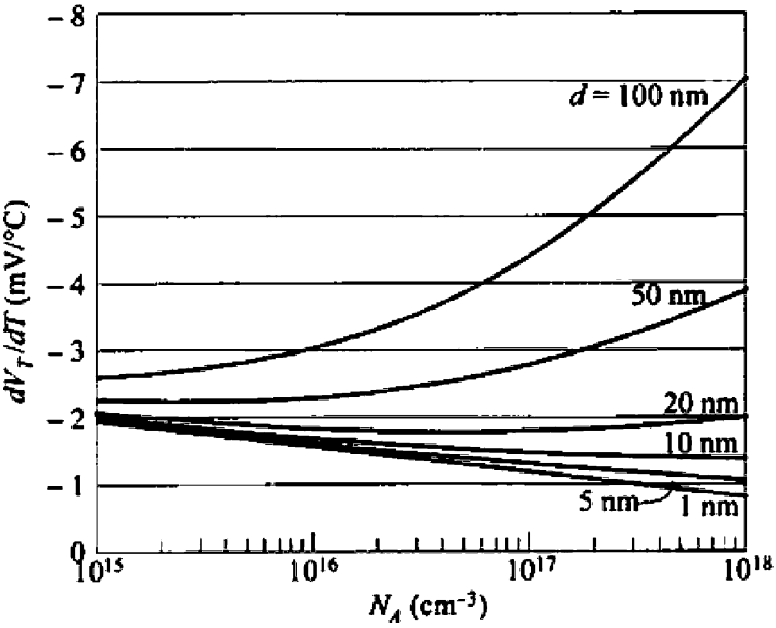

Fig. 2.4 Change in threshold voltage temperature dependence at room temperature vs. dopant concentration, with oxide thickness d [2]

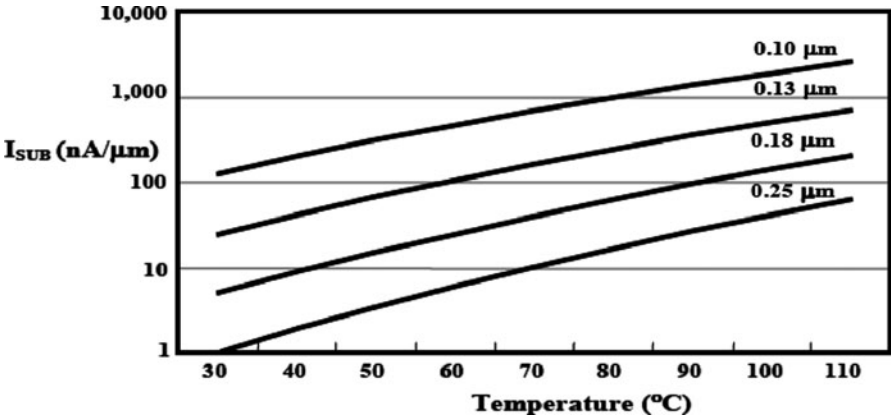

Fig. 2.5 Temperature dependence of subthreshold leakage current ($V_{GS} = 0$ V) [14]

$$I_0 = ATe^{-\frac{1.12}{2\phi_T}} \tag{2.16}$$

where I_0 is the reverse saturation current [12], A is a constant, and V_{DS} is the drain-source voltage. Recalling that $\phi_T = kT/q$, we see that I_0 is responsible for the exponential temperature dependence shown in Fig. 2.5.

The temperature dependence of gate leakage current has been shown to be very minor compared to that of subthreshold leakage current [13].

2.1.9 Interconnect Resistance

The interconnect resistance R is related to temperature by

$$R(T) = R_0[1 + \alpha_R(T - T_0)] \tag{2.17}$$

where T is the temperature, R_0 is the resistance at nominal temperature T_0, and α_R is an empirical term named the temperature coefficient of resistance. Al and Cu interconnects have similar values of α_R—0.004308 and 0.00401, respectively. Over the military-specified temperature range, Al wire resistances can change by up to 77.5% while Cu wire resistances can change by up to 72.2%. Interconnect resistance increases with increasing temperature, complicating evaluation of the impact of temperature on interconnect links—in these applications, the MOSFET currents may either increase or decrease in temperature (as explored in the next subsection), which means that the impact of temperature on interconnect resistance can either add to the system temperature dependence or reduce the temperature dependence, depending on the operating conditions.

2.1.10 Electromigration

Electromigration is a failure mechanism caused by high-energy electrons impacting the atoms in a material and causing them to shift position. It is most problematic in areas of high current density. This can form a positive feedback path can form where electromigration will cause an atom to move down a wire, slightly narrowing the wire width at that location and increasing the current density; this increased current density then further increases electromigration, causing more atoms to be displaced. This brings about two failure mechanisms: (1) the narrowing of the wire will increase wire resistance, which may cause a timing failure if a signal can no longer propagate within the clock period, or (2) electromigration will continue until the wire completely breaks, allowing no further current flow and resulting in functional failure.

Electromigration's impact on a system's reliability is measured in terms of a mean time to failure (MTTF) using Black's equation [15]

$$MTTF = A_j \cdot J^{-n_j} \cdot e^{\frac{E_a}{kT}} \tag{2.18}$$

where A_j is a constant related to the cross-sectional area of a wire, J is the current density, n is a constant scaling factor, E_a is the activation energy, k is the Boltzmann constant, and T is temperature. Thus, the MTTF is exponentially dependent on temperature.

2.2 Normal and Reverse Temperature Dependence

Changes in temperature affect system speed, power, and reliability by altering the threshold voltage [11], mobility [11], and saturation velocity [16] in each device. The resulting changes in device current can lead to failures in timing, cause systems to exceed power or energy budgets, and result in communication errors between IP cores. The temperature relationships for MOSFET mobility, threshold voltage, and velocity saturation are related to temperature using the following empirical expressions [17]:

$$\mu(T) = \mu_0 (T/T_0)^{\alpha_\mu} \tag{2.19}$$

$$V_T(T) = V_{T0} + \alpha_{V_T}(T - T_0) \tag{2.20}$$

$$v_{sat}(T) = v_{sat0} + \alpha_{v_{sat}}(T - T_0) \tag{2.21}$$

where T is the temperature; T_0 is the nominal temperature; μ_0, V_{T0}, and v_{sat0} are the mobility, threshold voltage, and saturation velocity at T_0, respectively; α_μ, α_{VT}, and α_{vsat} are empirical parameters named the mobility temperature exponent, threshold voltage temperature coefficient, and saturation velocity temperature coefficient, respectively, where $\alpha_\mu \approx -1.3$, $\alpha_{VT} \approx -3$ mV/°C, and $\alpha_{vsat} \approx -97$ m/(s·°C). Two temperature dependencies exist: the normal dependence (ND) region, where drain current (I_D) *decreases* with increasing temperature, and the reverse dependence (RD) region, where I_D *increases* with increasing temperature [18]. Between the two regions, there is a supply voltage where the impact of temperature on delay is minimized. This is referred to as the temperature-insensitive voltage V_{INS} [19], and as technology scales this voltage approaches nominal voltage.

In the temperature region of concern (between −55°C and 125°C, the range of military operating temperatures [20]), μ, V_T, and v_{sat} all decrease with increasing temperature. Examining the velocity-saturated MOSFET drain current $I_D(T)$ [21] we see that decreasing v_{sat} *decreases* I_D, while decreasing V_T *increases* I_D [22].

$$I_D(T) = v_{sat}(T) \cdot W \cdot P_s \cdot [V_{GS} - V_T(T)]^{\alpha} \tag{2.22}$$

Where W is the device width, P_s is a technology-specific constant, V_{GS} is the MOSFET gate-source voltage, and α is a technology-specific exponent. The temperature dependence of the MOSFET drain current, dI_D/dT, can be determined by the sum of the impacts of v_{sat} and V_T on I_D, composed of four values—(1) the change in velocity

saturation for a change in temperature, dv_{sat}/dT, (2) the change in threshold voltage for a change in temperature, dV_T/dT, (3) the change in device current for a change in velocity saturation, $\partial I_D/\partial v_{sat}$, and (4) the change in device current for a change in threshold voltage, $\partial I_D/\partial V_T$:

$$\left.\frac{dI_D}{dT}\right|_{Tot} = \left.\frac{dI_D}{dT}\right|_{v_{sat}} + \left.\frac{dI_D}{dT}\right|_{v_T} = \frac{\partial I_D}{\partial v_{sat}} \cdot \frac{dv_{sat}}{dT} + \frac{\partial I_D}{\partial v_T} \cdot \frac{dv_T}{dT} \tag{2.23}$$

$dI_D/dT|_{vsat}$ is negative, and $dI_D/dT|_{VT}$ is positive. At nominal voltage in conventional CMOS technologies, the magnitude of $dI_D/dT|_{vsat}$ is greater than the magnitude of $dI_D/dT|_{VT}$; thus, circuits at nominal voltages become slower as temperature increases. However, as V_{GS} approaches V_T, a change in V_T has a larger impact on I_D; thus, at lower supply voltages, the magnitude of $dI_D/dT|_{vsat}$ is less than the magnitude of $dI_D/dT|_{VT}$, and device delay decreases as temperature increases (the reverse temperature dependence). V_{INS} occurs where $dI_D/dT|_{Tot}$ approaches zero, with $dI_D/dT|_{vsat} \approx -dI_D/dT|_{VT}$; however, because v_{sat} and V_T differ between NMOS and PMOS devices, each type of device has a different value of V_{INS}. The dependence regions are shown in Fig. 2.6 for plots of the current through diode-connected PMOS and NMOS devices in a 90 nm technology model [23] over the range of military operating temperatures. In Fig. 2.6a, V_{INS} occurs in the shaded regions, with higher voltages exhibiting the normal temperature dependence and lower voltages exhibiting the reverse temperature dependence.

The reverse temperature dependence is occasionally referred to as temperature inversion, while the normal and reverse temperature dependences are also referred to as negative (for normal dependence) and positive (for reverse dependence) current-temperature (*I-T*) slopes. In this document, we will use the $\pm$ *I-T* slope terminology as shorthand for the normal and reverse temperature dependences.

The difference between the 125°C and −55°C endpoints of Fig. 2.6a is presented in Fig. 2.6b. In Fig. 2.6b, V_{INS} is indicated in each device by the minimum points in each curve; the absolute minimum for a 1:1 sizing ratio occurs at 345 mV, corresponding to an 18% total change in current over the entire 180°C range of ambient temperatures.

2.2.1 *Discovery of the Normal and Reverse Temperature Dependences*

This book is by no means the first document to report on the reverse temperature dependence. Indeed, what we name the reverse temperature dependence (i.e. the increasing of electrical conduction with increasing temperature) was first discovered by Faraday with his silver sulphide experiments mentioned in the previous chapter. However, the mechanism detected by Faraday was quite different than the

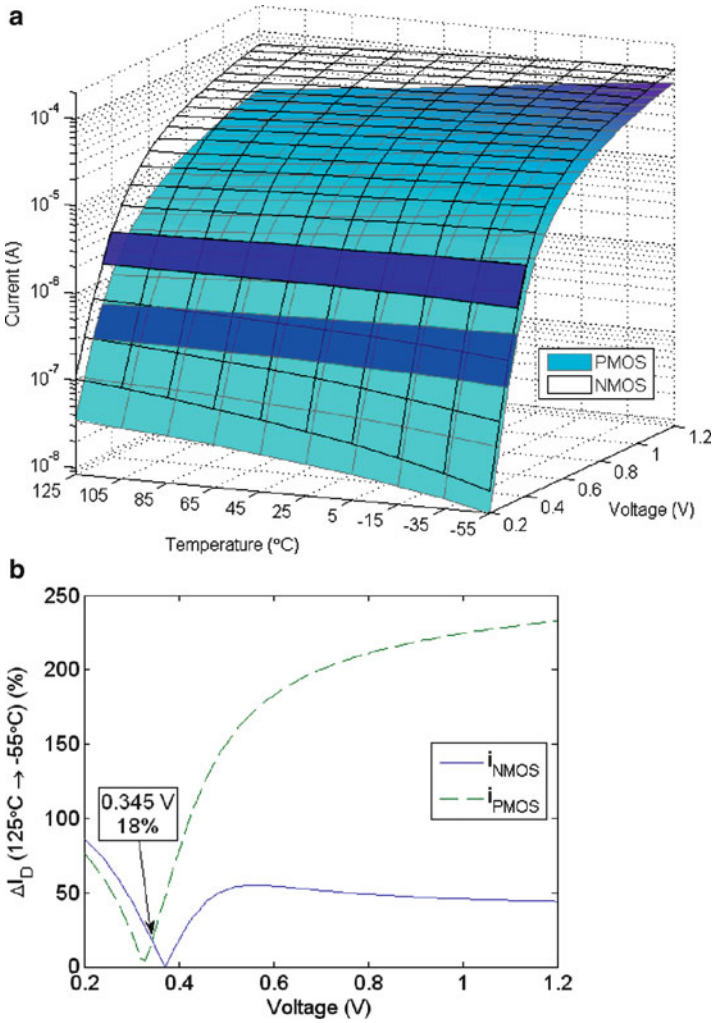

Fig. 2.6 (**a**) Device current across a range of temperatures and supply voltages in a 90 nm technology, (**b**) temperature change from 125°C to −55°C

mechanisms causing the normal and reverse temperature dependences in our modern Silicon electronics. The first recorded mention of the reversal of the temperature dependence describing the trade-off between mobility and threshold voltage is attributed to C. Park, *et al.*, from Motorola in 1995, in a conference paper exploring the impact of temperature on integrated circuits at very low voltages [18]. In the time since, the reversal of the temperature dependence has been explored in great detail [11, 24, 25], including magazines, patents, and journal papers, and is now being considered in industry-standard tools [26].

2.3 Temperature and Technology Scaling

V_{INS} occurs at increasingly larger supply voltages as technology is scaled, and is fast approaching the nominal supply voltage V_{NOM} (which is steadily decreasing as technology scales) as shown in Table 2.2, particularly with the introduction of high-κ dielectrics to replace SiO_2 [27] (high-κ dielectrics reduce μ and change $\partial\mu/\partial T$ [28], altering the balance of the μ and V_T impacts). The change in V_{INS} as technology scales is caused by changes in the threshold voltage and saturation velocity [29]. The data in Table 2.2 were generated using predictive technology models [23], with the 45, 32, and 22 nm data points using high-κ dielectric/metal gate models.

As V_{INS} approaches V_{NOM}, adaptive systems which vary supply voltage to reduce energy consumption or improve reliability will have operating voltages in both the normal and reverse temperature dependence regions, making it unclear if circuits will increase in speed or decrease in speed as chip temperatures increase, and exacerbating problems associated with inter-die temperature variation; solutions to this issue are discussed in Chap. 3.

The current dependencies for a 22 nm technology with high-κ dielectrics and metal gates are shown in Fig. 2.7. As shown, V_{INS} in the PMOS device increases from ~375 mV at 90 nm to ~575 mV at 22 nm, with the nominal supply voltage decreasing from 1.2 V at 90 nm to 0.8 V at 22 nm. The 22 nm NMOS device is in the reverse temperature dependence region even at the nominal supply voltage.

2.3.1 The Reverse Temperature Effect and High-κ/Metal Gate Technology

V_T, μ, v_{sat} and nominal supply voltage are all technology dependent parameters, with predicted values available down to the 22 nm node [23, 30]. Use of high-κ dielectrics and metal gates to alleviate nanoscale gate leakage problems also alters V_T, μ, and v_{sat} [31, 32]. The combination of these changes makes it difficult to determine the effect of temperature on device performance. Two dependences exist, as mentioned in the prior subsection: a normal temperature dependence, where current decreases as temperature increases, and a reverse temperature dependence [18, 19], where current increases as temperature increases.

Table 2.2 V_{INS} approaches V_{NOM} as technology scales

Technology (nm)	V_{NOM} (V)	V_{INS} (V)	V_{INS}/V_{NOM}
90	1.2	0.39	0.33
65	1.1	0.40	0.36
45	1.0	0.61	0.61
32	0.9	0.69	0.77
22	0.8	0.73	0.91

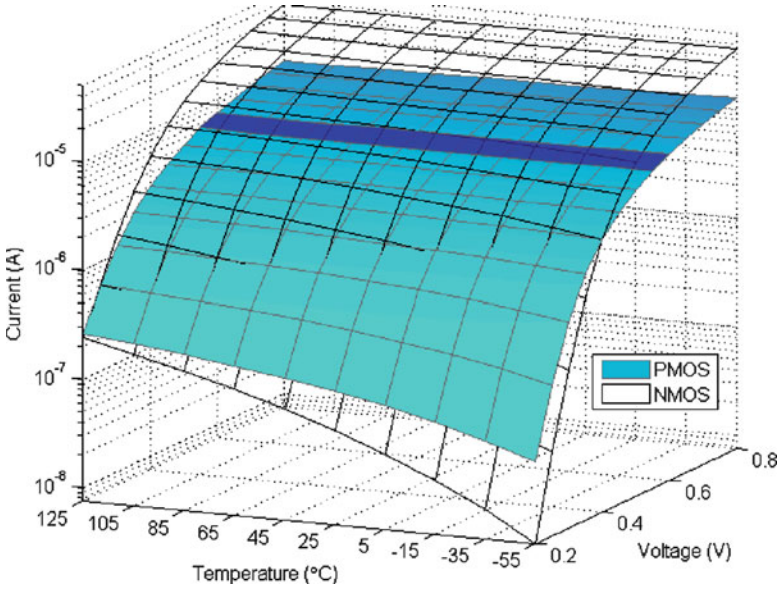

Fig. 2.7 Temperature dependence of device current across a range of supply voltages in a 22 nm high-κ/metal gate technology

These parameters are further complicated by environmental requirements (military specifications call for a range of −55–125°C [20]) and intra-die temperature variation (shown to exceed 50°C [33]). To account for the wide range of conditions, as well as process and voltage variations, variation-tolerant adaptive systems have been used to guarantee functionality by adjusting operating voltages and frequencies [34–36]; however, these systems with multiple voltage modes make the above-mentioned temperature effects even more difficult to determine.

For large gate overdrives ($V_{GS}-V_T > V_{INS}$), the temperature dependence of a device is dominated by the dependence of v_{sat}, while for small gate overdrives ($V_{GS}-V_T < V_{INS}$), small changes in V_T can cause large changes in current, resulting in a temperature dependence dominated by V_T. Further examination of these effects in SiO_2 dielectric, polysilicon gate devices is available in [18, 19].

In nanoscale devices, high-κ dielectrics and metal gates have been introduced to reduce gate leakage due to thinning gate oxides and reduce the depletion effects of polysilicon gates [31, 32]; unfortunately, these techniques have the effect of dramatically altering the temperature dependence of I_D. The extent of the change is shown in Fig. 2.8, which compares 45 nm predictive technology models [23] of both SiO_2/poly gate (dashed line) and high-κ/metal gate (solid line) devices. Each line in Fig. 2.8 shows the change in delay of an inverter (sizing ratio $\beta = 2$) from −55°C to 125°C. For example, at 0.62 V, the high-κ/metal gate inverter delay is unchanged from −55°C to 125°C, resulting in the 0.62 V point occurring on the 0% line. This 0% intersect point on each curve represents V_{INS}. As shown, V_{INS} in the high-κ/metal gate is 40% higher than in the SiO_2/poly gate devices.

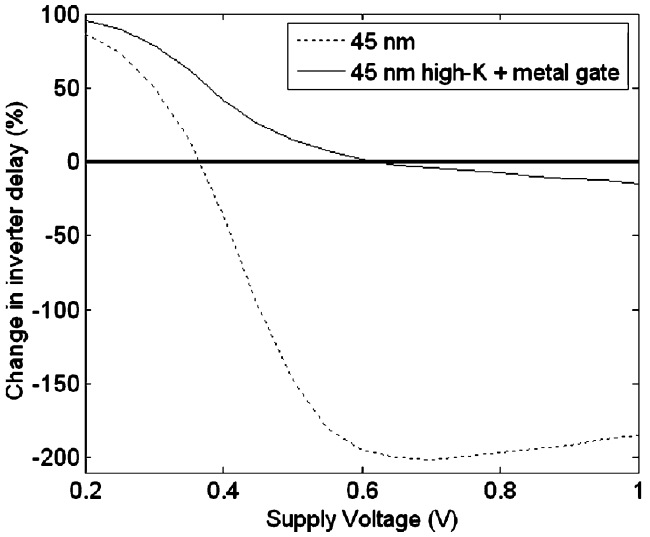

Fig. 2.8 Effect of high-κ dielectric and metal gate on temperature dependence

The normal dependence region is below the 0% line, and the reverse dependence region is above the 0% line.

Fig. 2.9a shows the change in PMOS device current from −55°C to 125°C at the 45, 32, and 22 nm technology nodes (with nominal voltages of 1, 0.9, and 0.8 V, respectively). As shown, V_{INS} increases by ~40 mV per technology node, with V_{INS} at 22 nm equal to 0.56 V. The NMOS device response, shown in Fig. 2.9b, is in the reverse temperature dependence region over the entire range of operating voltages at the 32 and 22 nm nodes.

The PMOS and NMOS devices are combined into an inverter with $\beta = 2$ in Fig. 2.9c. As shown, V_{INS} approaches 90% of nominal voltage in the 22 nm node. As β increases, the stronger PMOS effect decreases V_{INS}. Thus, adaptive voltage systems may easily wind up straddling both temperature domains in nanoscale systems, making temperature-aware design increasingly critical as technology scales.

Reverse temperature dependence at near nominal voltages complicates variation-tolerant system design, which uses multiple supply voltages to adjust for changes in process, voltage, and temperature. The additional complexity needed to account for both normal and reverse temperature dependence depends on the available design time information. If the system can be fully characterized at design time, then the multiple dependences can be programmed into the voltage and frequency look-up table entries [34] to ensure that the system adapts in the correct direction given a change in temperature. For example, whereas a low-voltage system would generally reduce the frequency as temperature increases, in the reverse dependence region the system would have to reduce the frequency when temperature decreases.

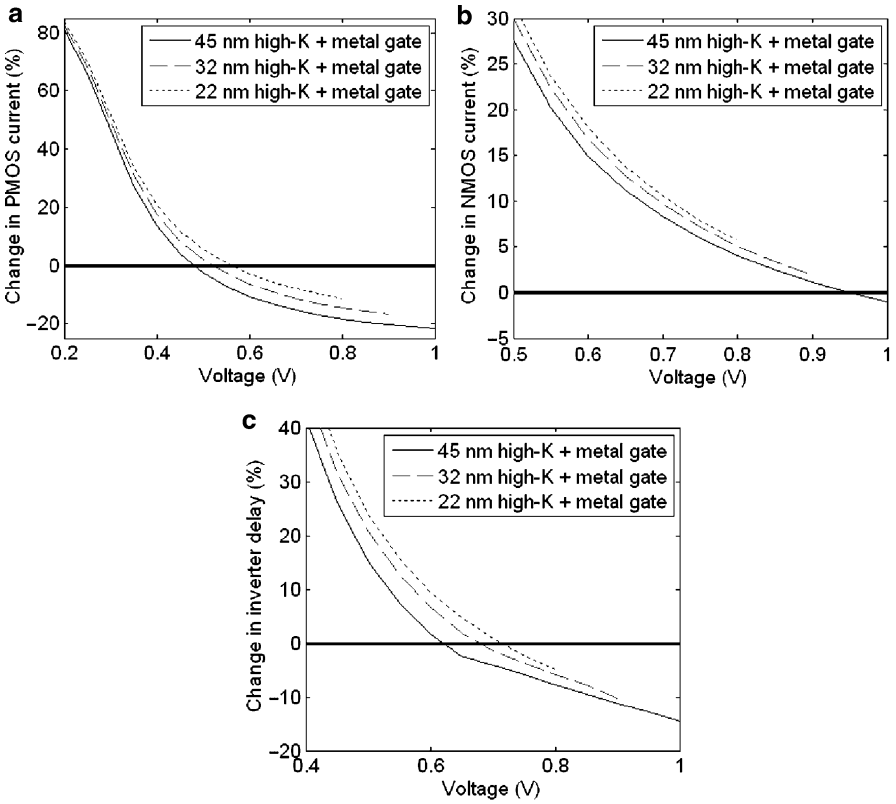

Fig. 2.9 Changes in (**a**) PMOS current, (**b**) NMOS current, and (**c**) inverter delay over the −55°C to 125°C temperature range

The reverse temperature effect is particularly important to consider in adaptive systems because of thermal runaway. In the normal dependence region, temperatures are prevented from increasing to dangerous levels because the delay becomes so large that the adaptive system is forced to reduce the clock frequency, reducing the energy and therefore the temperature. In the reverse temperature dependence region, circuits continue to speed up as temperature increases; there is no delay limit on high temperature operation. The higher temperatures could result in thermal runaway resulting from the exponential temperature dependence of leakage current [37], which may already be dominating the total power consumption in the nanoscale regime [38].

If the temperature dependences are not known at design time (from tool limitations, process variations, unknown IR drops, etc.), there are two options to ensure variation-tolerance. The system may be designed with large enough guardbands that it can operate correctly over the entire temperature range regardless of the temperature dependence, though this will result in a large reduction in delay performance. Another option is to use a temperature dependence sensor to determine the temperature dependence at each operating voltage; we propose the first temperature dependence sensor in Chap. 3.

References

1. Varshni YP (1967) Temperature dependence of the energy gap in semiconductors. Physica 34:149–154
2. Sze SM (1981) Physics of semiconductor devices, 2nd ed. John Wiley and Sons, NY
3. Chain K, Huang JH, Duster J, Ko PK, Hu C (1997) A MOSFET electron mobility model of wide temperature range (77–400K) for IC simulation. Semicond Sci Technol 12:355–358
4. Sabnis AG, Clemens JT (1979) Characterization of the electron mobility in the inverter Si surface. Int Electron Devices Mtg 18–21
5. Chen K, Wann HC, Dunster J, Ko PK, Hu C (1996) MOSFET carrier mobility model based on gate oxide thickness, threshold and gate voltages. Solid-State Electronics 39:1515–1518
6. Jeon DS, Burk DE (1989) MOSFET electron inversion layer mobilities–a physically based semi-empirical model for a wide temperature range. IEEE Trans Electron Devices 36:1456–1463
7. Grabinski W, Bucher M, Sallese JM, Krummenacher F (2000) Compact modeling of ultra deep submicron CMOS devices. Int Conf on Signals and Electronic Systems 13–27
8. Fang FF, Fowler AB (1970) Hot electron effects and saturation velocities in Silicon inversion layers. J Appl Phys 41:1825–1831
9. Cheng Y et al (1997) Modelling temperature effects of quarter micrometer MOSFETs in BSIM3v3 for circuit simulation. Semicond Sci Technol 12:1349–1354
10. Pierret RF (1988) Semiconductor fundamentals, 2nd ed. Addison-Wesley, MA
11. Filanovsky IM, Allam A (2001) Mutual compensation of mobility and threshold voltage temperature effects with applications in CMOS circuits. IEEE Trans Circuits and Syst I: Fundamental Theory and Applications 48:876–884
12. Oxner ES (1988) FET technology and application. CRC Press, NY
13. Agarwal A, Mukhopadhyay S, Raychowdhury A, Roy K, Kim CH (2006) Leakage power analysis and reduction for nanoscale circuits. IEEE Micro 26:68–80
14. Fallah F, Pedram M (2005) Standby and active leakage current control and minimization in CMOS VLSI systems. IEICE Trans Electronics E88-C:509–519
15. Black JR (1969) Electromigration–a brief survey and some recent results. IEEE Trans Electron Devices 16:338–347
16. Ku JC, Ismail Y (2007) On the scaling of temperature-dependent effects. IEEE Trans Computer-Aided Design of Integrated Circuits and Syst 26:1882–1888
17. Morshed TH et al (2009) BSIM4.6.4 MOSFET model user's manual. [Online] http://www-device.eecs.berkeley.edu/~bsim3/bsim4_arch_ftp.html
18. Park C et al (1995) Reversal of temperature dependence of integrated circuits operating at very low voltages. Int Electron Devices Mtg 71–74
19. Bellaouar A, Fridi A, Elmasry MI, Itoh K (1998) Supply voltage scaling for temperature-insensitive CMOS circuit operation. IEEE Trans Circuits and Syst II: Analog and Digital Signal Processing 45:415–417
20. US Dept of Defense (2007) Integrated circuits (microcircuits) manufacturing, general specification, Std MIL-PRF-38535H. Washington DC
21. Sakurai T, Newton AR (1990) Alpha-power law MOSFET model and its applications to CMOS inverter delay and other formulas. IEEE J Solid-State Circuits 25:584–594
22. Shichman H, Hodges DA (1968) Modeling and simulation of insulated-gate field-effect transistor switching circuits. IEEE J Solid-State Circuits SC-3:285–289
23. Zhao W, Cao Y (2006) New generation of predictive technology model for sub-45nm early design exploration. IEEE Trans Electron Devices 53:2816–2823
24. Lasbouygues B, Wilson R, Azemard N, Maurine P (2006) Timing analysis in presence of supply voltage and temperature variations. ACM Int Symp on Physical Design 10–16
25. Kumar R, Kursun V (2006) Reversed temperature-dependent propagation delay characteristics in nanometer CMOS circuits. IEEE Trans Circuits and Syst II: Express Briefs 53:1078–1082

26. Dasdan A, Hom I (2006) Handling inverted temperature dependence in static timing analysis. ACM Trans Design Automation of Electronic Syst 11:306–324
27. Wolpert D, Ampadu P (2008) Normal and reverse temperature dependence in variation-tolerant nanoscale systems with high-k dielectrics and metal gates. 3rd ACM Int Conf on Nano-networks 1–5
28. Wong H, Iwai H (2006) On the scaling issues and high-k replacement of ultrathin gate dielectrics for nanoscale MOS transistors. Microelectronic Engineering 83:1867–1904
29. Langen D, Ruckert U (2002) Extending scaling theory by adequately considering velocity saturation. 15th Ann IEEE Int ASIC/SoC Conf 145–149
30. Zhao W (2008) Personal communication
31. Guillaumot B et al (2002) 75nm damascene metal gate and high-k integration for advanced CMOS devices. Int Electron Devices Mtg 355–358
32. Cheng B et al (1999) The impact of high-k gate dielectrics and metal gate electrodes on sub-100 nm MOSFETs. IEEE Trans Electron Devices 46:1537–1544
33. Sato T, Ichimiya J, Ono N, Hachiya K, Hashimoto M (2005) On-chip thermal gradient analysis and temperature flattening for SoC design. IEICE Trans Fundamentals E88-A:3382–3389
34. Tschanz J et al (2007) Adaptive frequency and biasing techniques for tolerance to dynamic temperature-voltage variations and aging. IEEE Int Solid-State Circuits Conf 292–604
35. Elgebaly M, Sachdev M (2007) Variation-aware adaptive voltage scaling system. IEEE Trans Very Large Scale Integr Syst 15:560–571
36. Martin S, Flautner K, Mudge T, Blaauw D (2002) Combined dynamic voltage scaling and adaptive body biasing for lower power microprocessors under dynamic workloads. IEEE/ACM Int Conf on Computer-Aided Design 721–725
37. Lee CC, de Groot J (2006) On the thermal stability margins of high-leakage current packaged devices. 8th Electronics Packaging Technology Conf 487–491
38. Kim NS et al (2003) Leakage current: Moore's law meets static power. Computer 36:68–75

http://www.springer.com/978-1-4614-0747-8

Managing Temperature Effects in Nanoscale Adaptive Systems

Wolpert, D.; Ampadu, P.

2012, XXII, 174 p., Hardcover
ISBN: 978-1-4614-0747-8